Certain Computational Aspects of Power Efficiency and of State Space Models

Richard Gagnon
Department of Mathematics
University of Maryland, College Park, MD 20742

March 18, 2005

Contents

1 Introduction

This chapter introduces the two research areas presented in this dissertation.

1.1 Computational Aspects of Power Efficiency

Fokianos, Kedem, Qin, Short (FKQS) (2001) [9] introduced a semiparametric approach to the one-way layout that relies on an exponential distortion between each of the m distributions associated with the m random samples. The classic approach to the one-way layout assumes that each of the m distributions are Gaussian with a common variance. Under the Gaussian assumption, the density ratios of the m distributions are exponential distortions of the form $g_i(x)/g_r(x) = \exp(\alpha_i + \beta_i x)$ for $i = 1, \ldots, m-1$ where one of the m distributions is chosen as the reference distribution $G_r(x)$ with density $g_r(x)$. The semiparametric approach generalizes the classic approach by generalizing the form of the density ratios to $g_i(x)/g_r(x) = \exp(\alpha_i + \beta_i h(x))$ for $i = 1, \ldots, m-1$ where $h(x)$ is chosen based on the application. The semiparametric approach utilizes a profile likelihood in order to develop maximum likelihood estimators for each of the distortion parameters $\{(\alpha_i, \beta_i) : i = 1, \ldots, m-1\}$ and for the reference distribution $G_r(x)$. The resulting semiparametric test evaluates the maximum likelihood estimator for β_i in order to test whether the unknown distortion parameter β_i equals zero; in other words, whether the two distributions are the same. The density ratios are examples of weight functions that depend on an unknown finite-dimensional parameter $\boldsymbol{\theta}$. Gilbert (2000) [12] examines the large sample theory of maximum likelihood estimates in semiparametric biased sampling models with respect to a common underlying distribution G. In that paper, Gilbert characterizes conditions, on the weight functions and on the random samples and their distributions, in order that $(\hat{\boldsymbol{\theta}}, \mathbb{G}_n)$ is uniformly consistent, asymptotically Gaussian, and efficient, where $\hat{\boldsymbol{\theta}}$ and $\mathbb{G}_n$ are the maximum likelihood estimators of $\boldsymbol{\theta}$ and G. As an example of this semiparametric approach, Qin and Zhang (1997) [21] tested the validity of logistic regression under case-control sampling with $m = 2$ and $h(x) = x$. More recently, [9] applied this semiparametric approach to rain-rate data from meteorological instruments. Simulation results in [9] have shown that the semiparametric test compares favorably with the common t-test.

A natural way to compare the semiparametric test and the t-test is to use the concepts of relative efficiency and Pitman efficiency [2]. Relative efficiency is the ratio of the sample sizes for each test needed to achieve a desired power when the m distributions are different. The limit of the relative efficiency as each of the $m-1$ distorted distributions converge to the reference distribution in a prescribed manner is called the Pitman efficiency. This chapter presents as original work an analysis of the relative and Pitman efficiency of the semiparametric test versus the common t-test when there are $m = 2$ distributions. As part of this analysis, the generalized Glivenko-Cantelli theorem from [30] and the theory of extremum estimators from [1] are used to find asymptotic Gaussian test distributions of the semiparametric test and the t-test under the alternative hypothesis that the two random sample distributions are different. The asymptotic Gaussian test distributions are found for four examples of the random sample distributions: a Gaussian example, two gamma examples, and a log normal example. An efficiency analysis is then developed that establishes a theoretical efficiency based on Gaussian test distributions. The asymptotic Gaussian test distributions for each of the four examples are then applied to find the corresponding theoretical efficiency. Simulation results are then reported that verify the theoretical results for each example of the random sample distributions. For the Gaussian example, the efficiency of the semiparametric test versus the t-test is very close to one when the distortion parameter β is close to zero. For the other three examples, the semiparametric test is more efficient than the t-test for large parameter ranges of the random sample distributions.

1.2 Computational Aspects of State Space Models

Linear state space models provide a methodology for studying time series in discrete time [3], [7], [10], [13], [14], [17], [26], [29]. A large class of linear state space models provide a way to formalize the relationship between an unobservable time series (consisting of unknown states) and an observable time series. R. E. Kalman (1960) [13] introduced the Kalman filter as a sequential algorithm that provides a predictor (one step ahead) estimate and a filter estimate of each state based on the available observations at each time point under a Gaussian assumption, see also [3], [7], [10], [14], [17], [26], [29]. As part of the Kalman predictor and filter, variances (called precisions) are provided of the residuals between each state and its predictor and filter estimates. An important extension to the Kalman filter was the development of the state space smoother by Rauch (1962) [24] and by Bryson and Frazier (1963) [5], see also Rauch, Tung, and Striebel (1965) [25]. The state space smoother provides smoother estimates of all existing or past states as new or future observations become available [7], [10], [14], [17], [29]. Precisions of the smoother residuals are also provided. The state space smoother has several equivalent forms [7], [10], that include: the fixed interval smoother, the fixed point smoother, and the fixed lag smoother. Asymptotic analysis has shown that the precision of the Kalman filter estimate of the state associated with the most recent observation converges to a steady state value under certain conditions [7], [10], [29].

Under the Gaussian assumption the Kalman estimates of each state and precisions are the conditional means of each state and conditional error covariances given the available observations. These Kalman estimates of each state are optimal in the sense that the associated precisions are the minimum possible within the class of state estimators given the available observations. It turns out that the Kalman equations still hold when the Gaussian assumption is removed. In this case, the Kalman estimates of each state are the projection of each state on the subspace spanned by the available observations and the precisions are the minimum least square error estimators within the class of linear state estimators, see section 4.2 and problems 4.4 and 4.6 in [29] and section 12.2 in [4]. In this case, these Kalman estimates of each state are suboptimal in the sense that the resulting precisions are larger that the precisions associated with the true conditional mean of each state given the available observations.

This chapter provides as original work an analysis of the smoother precisions where the observable and unobservable time series are univariate and where the state space parameters are constant. This analysis starts by introducing a likelihood smoother form of the state space smoother based on a general multivariate version of the linear Gaussian state space model. This analysis then applies the likelihood smoother to a univariate version of the linear Gaussian state space model with constant parameters in order to develop a variety of upper and lower bounds on the smoother precisions and also to develop the asymptotic behavior of the smoother precisions as the number of observations increases. These asymptotic smoother precision values provide a way to evaluate the future evolution of the smoother precision values associated with a finite time series as new observations become available. This chapter concludes by introducing the partial (suboptimal) state space smoother that provides a smoother like estimate of each state that only relies on a limited number of future observations.

2 Computational Aspects of Power Efficiency

In this chapter the relative efficiency of the semiparametric test versus the common t-test is investigated. Section 2.1 summarizes some of the published mathematical theory behind the semiparametric approach. Section 2.1.1 identifies four examples of random sample distributions that are analyzed in detail throughout this chapter. Section 2.2 extends the current theory behind the semiparametric approach by developing a relative efficiency analysis of the semiparametric test versus the t-test. Section 2.2.1 develops an asymptotic Gaussian distribution for the semiparametric test under the alternative hypothesis that the two random sample distributions are different. An asymptotic distribution for the semiparametric test is found using each of the random sample examples identified in subsection 2.1.1. Section 2.2.2 also develops an asymptotic Gaussian distribution for the t-test under the alternative hypothesis that the two random sample distributions are different. An asymptotic distribution for the t-test is found using each of the random sample examples identified in section 2.1.1. Section 2.2.3 develops a relative efficiency analysis of the semiparametric test and the t-test given their asymptotic Gaussian test distributions. This section develops a relative efficiency using each of the random sample examples identified in subsection 2.1.1. In order to complement the relative efficiency theory, this section also contains a simulation study that supports the theoretical results for each of the random sample examples in subsection 2.1.1.

2.1 Some Preliminary Statistical Formulations

This section briefly presents the formulation of the semiparametric approach from [9] that is developed further in subsequent sections.

The classical one-way analysis of variance with $m = q + 1$ independent random samples is described as follows:

$$\begin{aligned} x_{11}, \ldots, x_{1n_1} &\sim X_1 \text{ with pdf } g_1(x) \\ &\vdots \\ x_{q1}, \ldots, x_{qn_q} &\sim X_q \text{ with pdf } g_q(x) \\ x_{m1}, \ldots, x_{mn_m} &\sim X_m \text{ with pdf } g_m(x) \end{aligned}$$

where $g_m(x)$ is arbitrarily labeled as the reference probability density, and where $g_j(x)$ is a probability density with finite mean and variance: $(\mu_j, \sigma_j^2), j = 1, \ldots, m$. Assuming that each of the m probability densities is Gaussian with common variance ($\sigma_1^2 = \cdots = \sigma_m^2 = \sigma^2$) implies an exponential distortion for each of the first q distributions, relative to the mth distribution, of the form

$$\frac{g_j(x)}{g_m(x)} = \exp(\alpha_j + \beta_j x),\ j = 1, \ldots, q \tag{1}$$
$$\alpha_j = \frac{\mu_m^2 - \mu_j^2}{2\sigma^2},\ \beta_j = \frac{\mu_j - \mu_m}{\sigma^2},\ j = 1, \ldots, q.$$

The semiparametric approach generalizes the analysis of the one-way layout by dropping the Gaussian probability density assumption and by generalizing the form of the exponential distortion:

$$w_j(x|\alpha_j, \beta_j) \equiv \frac{g_j(x)}{g_m(x)} = \exp(\alpha_j + \beta_j h(x)),\ j = 1, \ldots, q \tag{2}$$
$$w_m(x|\alpha_m, \beta_m) \equiv 1,\ (\alpha_m, \beta_m) \equiv \mathbf{0}$$

where $h(x)$ may assume various forms as shown in several examples below. Various generalizations of (2) have been suggested by Gilbert, Lele, and Vardi (1999) [11], and by Qin (1998) [20]. Observe that (2) is a special case of a weighted distribution as defined by Patil and Rao (1977) [19].

Let $\boldsymbol{x}_j = (x_{11}, \ldots, x_{1n_1})'$ identify the random sample from the jth probability density, for $j = 1, \ldots, m$; let $\boldsymbol{t} \equiv (t_1, \ldots, t_n)' = (\boldsymbol{x}_1', \ldots, \boldsymbol{x}_m')'$ identify the combined data from each of the m probability densities where $n = n_1 + \cdots + n_m$ identifies the combined sample size; let $\rho_j = n_j/n_m, j = 1, \ldots, m$ denote the sample proportions; and let $g(x) = g_m(x)$ identify the reference density. Then the semiparametric approach finds a maximum likelihood estimator for $G(x)$ (the cdf of $g(x)$) over the class of step cdf's with jumps at the observed values $t_i \in \boldsymbol{t}$.

With $p(t_i) = dG(t_i), i = 1, \ldots, n$ and $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \equiv ((\alpha_1, \ldots, \alpha_q), (\beta_1, \ldots, \beta_q))' \in \mathbb{R}^{2q}$, the likelihood becomes,

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, G) = \prod_{i=1}^{n} p(t_i) \prod_{j=1}^{n_1} \exp(\alpha_1 + \beta_1 h(x_{1j})) \cdots \prod_{j=1}^{n_q} \exp(\alpha_q + \beta_q h(x_{qj})) \tag{3}$$

Fixing $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and then maximizing (3) with respect to $p(t_i)$, subject to m constraints that the $p(t_i)$ and each of the distortions sum to 1,

$$\sum_{i=1}^{n} p(t_i) = 1,\ \sum_{i=1}^{n} p(t_i) \left[w_j\left(t_i|\alpha_j, \beta_j\right) - 1\right] = 0,\ j = 1, \ldots, q$$

results in the following formulas for $p(t)$ and $g(t)$

$$\tilde{p}(t|\boldsymbol{\alpha},\boldsymbol{\beta}) = 1/\left[n + \lambda_1(w_1(t|\alpha_1,\beta_1) - 1) + \cdots + \lambda_q(w_q(t|\alpha_q,\beta_q) - 1)\right]$$
$$\tilde{G}(t|\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i=1}^{n} I(t_i \le t)\tilde{p}(t_i|\boldsymbol{\alpha},\boldsymbol{\beta})$$

where the Lagrange multipliers $\boldsymbol{\lambda} \equiv \{\lambda_1,\ldots,\lambda_q\} \equiv \boldsymbol{\lambda}(\boldsymbol{\alpha},\boldsymbol{\beta})$ depend on $(\boldsymbol{\alpha},\boldsymbol{\beta})$ since the m constraints must be satisfied and where $I(B)$ is the indicator of the event B. The resulting profile likelihood is $\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta},\tilde{G})$.

The estimates $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) = ((\hat{\alpha}_1,\ldots,\hat{\alpha}_q)',(\hat{\beta}_1,\ldots,\hat{\beta}_q)')$, for the true distortion parameters $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$, are solutions of the following score equations in terms of the profile likelihood $\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta},\tilde{G})$ (see [9]) for $j = 1,...,q$,

$$\begin{aligned} 0 = \frac{\partial}{\partial\alpha_j}\log\mathcal{L}\Big|_{(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})} &= n_j - \lambda_j \sum_{i=1}^{n} \tilde{p}(t_i|\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) w_j(t_i|\hat{\alpha}_j,\hat{\beta}_j) \\ 0 = \frac{\partial}{\partial\beta_j}\log\mathcal{L}\Big|_{(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})} &= \sum_{i=1}^{n_j} h(x_{ji}) - \lambda_j \sum_{i=1}^{n} h(t_i)\tilde{p}(t_i|\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) w_j(t_i|\hat{\alpha}_j,\hat{\beta}_j)\ . \end{aligned}$$

Hence the Lagrange multipliers take the form $\boldsymbol{\lambda}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) = \{n_1,\ldots,n_q\}$ in order to meet the m constraints. The resulting formulas for $p(t)$ and $g(t)$ with the Lagrange multipliers fixed at $\boldsymbol{\lambda} = \{n_1,\ldots,n_q\}$ are

$$\begin{aligned} \hat{p}(t|\boldsymbol{\alpha},\boldsymbol{\beta}) &= 1/\left(n_m D_q(t|\boldsymbol{\alpha},\boldsymbol{\beta})\right) \\ \hat{G}(t|\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum_{i=1}^{n} I(t_i \le t)\hat{p}(t_i|\boldsymbol{\alpha},\boldsymbol{\beta}) \\ D_q(t|\boldsymbol{\alpha},\boldsymbol{\beta}) &= 1 + \rho_1 w_1(t|\alpha_1,\beta_1) + \cdots + \rho_q w_q(t|\alpha_q,\beta_q)\ . \end{aligned}$$

Define the semiparametric log-likelihood as $l(\boldsymbol{\alpha},\boldsymbol{\beta}) \equiv \log\mathcal{L}(\boldsymbol{\alpha},\boldsymbol{\beta},\hat{G})$. The estimates $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ are also solutions of the score equations in terms of the semiparametric log-likelihood $l(\boldsymbol{\alpha},\boldsymbol{\beta})$. Under regularity conditions, the solutions $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ are consistent and asymptotically normal with mean $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$, and a $2q \times 2q$ covariance matrix $\boldsymbol{\Omega}/n$ (see [9])

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_{\boldsymbol{\alpha}_0} \\ Z_{\boldsymbol{\beta}_0}\end{pmatrix} \sim \mathrm{N}\left(\mathbf{0},\boldsymbol{\Omega}\right),\ \boldsymbol{\Omega} = \mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1} \tag{4}$$
$$\mathbf{V} \equiv \mathbf{Var}\left[\frac{1}{\sqrt{n}}\boldsymbol{\nabla} l\left(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0\right)\right],\ -\frac{1}{n}\boldsymbol{\nabla}\boldsymbol{\nabla}' l\left(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0\right) \xrightarrow{P} \mathbf{S}\ \text{ as } n \to \infty$$
$$\boldsymbol{\nabla} \equiv \left(\frac{\partial}{\partial\alpha_1},\ldots,\frac{\partial}{\partial\alpha_q},\ \frac{\partial}{\partial\beta_1},\ldots,\frac{\partial}{\partial\beta_q},\right)'\ .$$

For the general case ($q > 1, m = q + 1$), definitions for the matrices $\mathbf{S}$ and $\mathbf{V}$ that compose $\boldsymbol{\Omega}$ are found in [9]. For the case ($q = 1, m = 2$), Qin and Zhang (1997) [21] showed

$$
\begin{aligned}
\boldsymbol{\Omega} &= \frac{1+\rho_1}{\rho_1}\begin{bmatrix} A_0 & A_1 \\ A_1 & A_2 \end{bmatrix}^{-1} - \frac{(1+\rho_1)^2}{\rho_1}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\
A_k &= \mathrm{E}\left(\frac{X_1^k}{D_1\left(X_1|\alpha_0,\beta_0\right)}\right),\ k = 0, 1, 2\ .
\end{aligned} \tag{5}
$$

Under the null hypothesis that the m probability densities are the same, $\mathbf{H}_0 : \boldsymbol{\beta}_0 = \mathbf{0}$, the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ reduces as shown in [9]

$$
\begin{aligned}
\sqrt{n}\hat{\boldsymbol{\beta}} &\xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \frac{1}{\mathrm{Var}(h(X_m))}\boldsymbol{A}_{11}^{-1}\right) \\
\mathrm{Var}(h(X_m)) &= \int h^2(x)dG(x) - \left(\int h(x)dG(x)\right)^2 .
\end{aligned}
$$

For the case ($q = 1, m = 2$), $\boldsymbol{A}_{11} = \rho_1/(1+\rho_1)^2$ is a scalar as shown in [9], such that under $\mathbf{H}_0$:

$$
\begin{aligned}
\mathrm{Z}_n &\equiv \sqrt{n}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sqrt{\mathrm{Var}(h(X_m))}\hat{\beta} \xrightarrow{d} \mathrm{Z} \sim \mathrm{N}(0,1) \\
\text{or } \mathcal{X}_1 &\equiv n\frac{\rho_1}{(1+\rho_1)^2}\mathrm{Var}(h(X_m))\hat{\beta}^2 \xrightarrow{d} \chi^2_{(1)}
\end{aligned} \tag{6}
$$

and $\mathbf{H}_0$ is rejected for extreme values of Z_n or $\mathcal{X}_1$. Since $\mathrm{Var}(h(X_m))$ is generally unknown, $\mathrm{Var}(h(X_m))$ is estimated using:

$$
\widehat{\mathrm{Var}}(h(X_m)) \equiv \sum_{i=1}^{n} h^2(t_i)\hat{p}(t_i|\hat{\alpha},\hat{\beta}) - \left(\sum_{i=1}^{n} h(t_i)\hat{p}(t_i|\hat{\alpha},\hat{\beta})\right)^2
$$

so the actual semiparametric statistic is:

$$
\tilde{\mathrm{Z}}_n \equiv \sqrt{n}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sqrt{\widehat{\mathrm{Var}}(h(X_m))}\hat{\beta}\ .
$$

2.1.1 Some Distortion Examples

The previous section has already identified one weighted distribution example, namely a Gaussian example in (1). This section identifies other weighted distribution examples that are used throughout this chapter.

2.1.1.1 Gaussian Example The first example restates the Gaussian distribution example, where each of the m random variables X_j has a different mean parameter μ_j and has a common variance parameter σ^2.

$$
\begin{aligned}
X_j \sim g_j(x) &= \mathrm{N}\left(\mu_j, \sigma^2\right), \ j = 1 \ldots m \\
\mathrm{E}\left(X_j\right) &= \mu_j, \ \mathrm{Var}\left(X_j\right) = \sigma^2 \\
\mathrm{E}\left(X_j^2\right) &= \sigma^2 + \mu_j^2, \\
\mathrm{E}\left(X_j^3\right) &= 2\sigma^2\mu_j, \\
\mathrm{E}\left(X_j^4\right) &= 2\sigma^2\left(\sigma^2 + 2\mu_j^2\right) \\
w_j\left(x|\alpha_j, \beta_j\right) &= \frac{g_j(x)}{g_m(x)} = \exp\left(\alpha_j + \beta_j x\right), \ j = 1 \ldots q \\
\left(\alpha_j, \beta_j\right) &= \left(\frac{\mu_m^2 - \mu_j^2}{2\sigma^2}, \ \frac{\mu_j - \mu_m}{\sigma^2}\right), \ j = 1 \ldots q \\
h\left(X_j\right) &= X_j \sim \mathrm{N}\left(\mu_j, \sigma^2\right), \ j = 1 \ldots m
\end{aligned}
$$

2.1.1.2 Gamma Example I The second example identifies a gamma distribution example, where each of the m random variables X_j has a common shape parameter α_γ and has a different scale parameter $\beta_{\gamma j}$.

$$
\begin{aligned}
X_j \sim g_j(x) &= \mathrm{Gamma}\left(\alpha_\gamma, \beta_{\gamma j}\right), \ j = 1 \ldots m \\
\mathrm{E}\left(X_j\right) &= \alpha_\gamma \beta_{\gamma j}, \ \mathrm{Var}\left(X_j\right) = \alpha_\gamma \beta_{\gamma j}^2 \\
\mathrm{E}\left(X_j^k\right) &= \frac{\Gamma\left(\alpha_\gamma + k\right)}{\Gamma\left(\alpha_\gamma\right)} \beta_{\gamma j}^k, \ \text{for } k = 1, 2, \ldots \\
w_j\left(x|\alpha_j, \beta_j\right) &= \frac{g_j(x)}{g_m(x)} = \exp\left(\alpha_j + \beta_j x\right), \ j = 1 \ldots q \\
\left(\alpha_j, \beta_j\right) &= \left(\alpha_\gamma \log\left(\frac{\beta_{\gamma m}}{\beta_{\gamma j}}\right), \ \frac{1}{\beta_{\gamma m}} - \frac{1}{\beta_{\gamma j}}\right), \ j = 1 \ldots q \\
h\left(X_j\right) &= X_j \sim \mathrm{Gamma}\left(\alpha_\gamma, \beta_{\gamma j}\right), \ j = 1 \ldots m
\end{aligned}
$$

2.1.1.3 Gamma Example II The third example is again a gamma distribution example, where each of the m random variables X_j has a different shape parameter $\alpha_{\gamma j}$ and has a common scale parameter β_γ.

$$
\begin{aligned}
X_j \sim g_j(x) &= \text{Gamma}\left(\alpha_{\gamma j}, \beta_\gamma\right),\ j = 1 \ldots m \\
\text{E}\left(X_j\right) &= \alpha_{\gamma j}\beta_\gamma,\ \text{Var}\left(X_j\right) = \alpha_{\gamma j}\beta_\gamma^2 \\
\text{E}\left(X_j^k\right) &= \frac{\Gamma\left(\alpha_{\gamma j} + k\right)}{\Gamma\left(\alpha_{\gamma j}\right)}\beta_\gamma^k,\ k = 1, 2, \ldots \\
w_j\left(x|\alpha_j, \beta_j\right) &= \frac{g_j(x)}{g_m(x)} = \exp\left(\alpha_j + \beta_j \log(x)\right),\ j = 1 \ldots q \\
\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} &= \begin{pmatrix} \log \frac{\Gamma(\alpha_{\gamma m})}{\Gamma(\alpha_{\gamma j})} + \left(\alpha_{\gamma m} - \alpha_{\gamma j}\right) \log \beta_\gamma \\ \left(\alpha_{\gamma j} - \alpha_{\gamma m}\right) \end{pmatrix},\ j = 1 \ldots q \\
h\left(X_j\right) &= \log\left(X_j\right),\ j = 1 \ldots m \\
\text{M}_{\log(X_j)}\left(t\right) &= \frac{\Gamma\left(\alpha_{\gamma j} + t\right)}{\Gamma\left(\alpha_{\gamma j}\right)}\beta_\gamma^t,\ t > -\alpha_{\gamma j}
\end{aligned}
$$

2.1.1.4 Log Normal Example The fourth example identifies a log normal distribution example, where each of the m random variables X_j has a different μ_{lj} parameter and a common σ_l^2 parameter.

$$
\begin{aligned}
X_j \sim g_j(x) &= \text{LN}\left(\mu_{lj}, \sigma_l^2\right),\ j = 1 \ldots m \\
\text{E}\left(X_j\right) &= e^{\mu_{lj} + \sigma_l^2/2},\ \text{Var}\left(X_j\right) = e^{2\mu_{lj} + \sigma_l^2}\left(e^{\sigma_l^2} - 1\right) \\
\text{E}\left(X_j^k\right) &= e^{k\mu_{lj} + k^2\sigma_l^2/2},\ k = 1, 2, \ldots \\
w_j\left(x|\alpha_j, \beta_j\right) &= \frac{g_j(x)}{g_m(x)} = \exp\left(\alpha_j + \beta_j \log(x)\right),\ j = 1 \ldots q \\
\left(\alpha_j, \beta_j\right) &= \left(\frac{\mu_{lm}^2 - \mu_{lj}^2}{2\sigma_l^2},\ \frac{\mu_{lj} - \mu_{lm}}{\sigma_l^2}\right),\ j = 1 \ldots q \\
h\left(X_j\right) &= \log\left(X_j\right) \sim \text{N}\left(\mu_{lj}, \sigma_l^2\right),\ j = 1 \ldots m
\end{aligned}
$$

2.2 Efficiency Development

Throughout this section, usage of the term "T test" refers to the t-test. Experimental power comparisons between the $\tilde{\mathrm{Z}}_n$ and T tests have provided empirical evidence that the $\tilde{\mathrm{Z}}_n$ test compares favorably with the T test when the underlying probability densities are Gaussian, i.e. the two tests appear to have practically the same power over specific parameter ranges. When the underlying probability densities are not Gaussian, the power of the $\tilde{\mathrm{Z}}_n$ test appears in some cases to be greater than the power of the T test. This section quantifies the theoretical power relationship between the $\tilde{\mathrm{Z}}_n$ and T tests by examining the efficiency of the T test in relation to the $\tilde{\mathrm{Z}}_n$ test. To develop this efficiency, the asymptotic distributions for the $\tilde{\mathrm{Z}}_n$ and T test statistics are identified.

2.2.1 Asymptotic Distribution of the $\tilde{\mathbf{Z}}_n$ Statistic

In this section the asymptotic distribution of the $\tilde{\mathrm{Z}}_n$ statistic is examined for the case $(q = 1, m = 2)$, under the alternative hypothesis, $\mathbf{H}_1 : \beta_0 \neq 0$, where (α, β) renames the distortion parameters (α_1, β_1) and where the true distortion parameters of (α, β) are denoted as (α_0, β_0). This examination proceeds by expanding $\tilde{\mathrm{Z}}_n$, minus a suitable offset, into a linear combination of four random variables. The law of large numbers, the abstract Glivenko-Cantelli theorem, and the asymptotic properties of extremum estimators are applied to find the asymptotic limit for the coefficients of the random variables. The multivariate central limit theorem is applied to find the asymptotic joint distribution of the random variables. The asymptotic results, for the coefficients and for the random variables, are combined to find the asymptotic distribution for the modified $\tilde{\mathrm{Z}}_n$ statistic. The modified $\tilde{\mathrm{Z}}_n$ statistic is:

$$\begin{aligned}
\tilde{\mathrm{Z}}_n^* &\equiv \tilde{\mathrm{Z}}_n - \sqrt{n}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sigma_h\beta_0 = \sqrt{\frac{n_1 n_2}{n}}\left(\hat{\sigma}_h(\hat{\alpha},\hat{\beta})\hat{\beta} - \sigma_h\beta_0\right) \\
\hat{\sigma}_h^2(\hat{\alpha},\hat{\beta}) &\equiv \widehat{\mathrm{Var}}(h(X_2)) \equiv \hat{\mu}_{h^2}(\hat{\alpha},\hat{\beta}) - \left(\hat{\mu}_h(\hat{\alpha},\hat{\beta})\right)^2 \\
\sigma_h^2 &\equiv \mathrm{Var}\left(h(X_2)\right) \equiv \mu_{h^2} - \left(\mu_h\right)^2 \\
\hat{\mu}_{h^k}(\alpha,\beta) &= \sum_{i=1}^{n} h^k\left(t_i\right)\hat{p}\left(t_i|\alpha,\beta\right),\ \mu_{h^k} = \mathrm{E}\left(h^k(X_2)\right), k = 1, 2, \ldots
\end{aligned}$$

The $\tilde{Z}_n^*$ random variable expansion proceeds by deriving an alternate expression for $\tilde{Z}_n^*$ based on a Taylor series expansion for $\hat{\sigma}_h^2(\hat{\alpha},\hat{\beta})$ around (α_0,β_0)

$$
\begin{aligned}
&\sqrt{\frac{n_1 n_2}{n}}\left(\hat{\sigma}_h(\hat{\alpha},\hat{\beta})\hat{\beta}-\sigma_h\beta_0\right)\\
=&\sqrt{\frac{n_1 n_2}{n}}\begin{pmatrix}0, & 1\end{pmatrix}\begin{pmatrix}\hat{\alpha}-\alpha_0\\ \hat{\beta}-\beta_0\end{pmatrix}\hat{\sigma}_h(\hat{\alpha},\hat{\beta})\\
+&\sqrt{\frac{n_1 n_2}{n}}\left(\hat{\sigma}_h^2(\alpha_0,\beta_0)-\sigma_h^2+\boldsymbol{\nabla}'\hat{\sigma}_h^2(\alpha,\beta)\big|_{(\acute{\alpha},\acute{\beta})}\begin{pmatrix}\hat{\alpha}-\alpha_0\\ \hat{\beta}-\beta_0\end{pmatrix}\right)\frac{\beta_0}{\left(\hat{\sigma}_h(\hat{\alpha},\hat{\beta})+\sigma_h\right)}\\
=&\sqrt{\frac{n_1 n_2}{n}}\left(\frac{\hat{\sigma}_h^2(\alpha_0,\beta_0)-\sigma_h^2}{\hat{\sigma}_h(\hat{\alpha},\hat{\beta})+\sigma_h}\right)\beta_0+\sqrt{\frac{n_1 n_2}{n}}\frac{\mathbf{Q}_n}{\hat{\sigma}_h(\hat{\alpha},\hat{\beta})+\sigma_h}\begin{pmatrix}\hat{\alpha}-\alpha_0\\ \hat{\beta}-\beta_0\end{pmatrix}
\end{aligned}
$$

where the gradient $\boldsymbol{\nabla}\hat{\sigma}_h^2(\alpha,\beta)\in\mathbb{R}^2$ is a column vector, where $\mathbf{Q}_n\in\mathbb{R}^2$ is a row vector defined as follows

$$
\begin{aligned}
\mathbf{Q}_n &\equiv \mathbf{Q}_n\left((\hat{\alpha},\hat{\beta}),(\acute{\alpha},\acute{\beta})\right)\\
&\equiv \left(0, \quad \hat{\sigma}_h(\hat{\alpha},\hat{\beta})\left(\hat{\sigma}_h(\hat{\alpha},\hat{\beta})+\sigma_h\right)\right)+\beta_0\ \boldsymbol{\nabla}'\hat{\sigma}_h^2(\alpha,\beta)\big|_{(\acute{\alpha},\acute{\beta})}
\end{aligned} \tag{7}
$$

and where the mean value theorem shows that $(\acute{\alpha},\acute{\beta})$ satisfies

$$
\begin{aligned}
(\alpha_\lambda,\beta_\lambda) &= \lambda\left(\hat{\alpha},\hat{\beta}\right)+(1-\lambda)\left(\alpha_0,\beta_0\right),\ \lambda\in[0,1]\\
\left(\acute{\alpha},\acute{\beta}\right) &= \left(\alpha_{\acute{\lambda}},\beta_{\acute{\lambda}}\right)\ \text{ for some } \acute{\lambda}\in[0,1]\ .
\end{aligned} \tag{8}
$$

A Taylor series expansion of the score equation around (α_0,β_0) and the mean value Theorem 6.7 from Kress (1998) [16] provides an expression for $(\hat{\alpha}-\alpha_0,\hat{\beta}-\beta_0)$:

$$
\mathbf{0}=\boldsymbol{\nabla}l\left(\alpha,\beta\right)\big|_{(\hat{\alpha},\hat{\beta})}=\boldsymbol{\nabla}l\left(\alpha,\beta\right)\big|_{(\alpha_0,\beta_0)}+\int_0^1\boldsymbol{\nabla}\boldsymbol{\nabla}'l\left(\alpha_\lambda,\beta_\lambda\right)\begin{pmatrix}\hat{\alpha}-\alpha_0\\ \hat{\beta}-\beta_0\end{pmatrix}d\lambda
$$

where the gradient $\boldsymbol{\nabla}l(\alpha,\beta)\in\mathbb{R}^2$ is a column vector and the hessian $\boldsymbol{\nabla}\boldsymbol{\nabla}'l(\alpha,\beta)\in\mathbb{R}^{2\times 2}$ is a matrix that satisfies

$$
\begin{aligned}
\boldsymbol{\nabla}\boldsymbol{\nabla}'l\left(\alpha,\beta\right)\big|_{(\grave{\alpha},\grave{\beta})}\begin{pmatrix}\hat{\alpha}-\alpha_0\\ \hat{\beta}-\beta_0\end{pmatrix} &= \int_0^1\boldsymbol{\nabla}\boldsymbol{\nabla}'l\left(\alpha_\lambda,\beta_\lambda\right)\begin{pmatrix}\hat{\alpha}-\alpha_0\\ \hat{\beta}-\beta_0\end{pmatrix}d\lambda\\
\left(\grave{\alpha},\grave{\beta}\right) &= \left(\alpha_{\grave{\lambda}},\beta_{\grave{\lambda}}\right)\ \text{ for some } \grave{\lambda}\in[0,1]\ .
\end{aligned} \tag{9}
$$

The resulting $\tilde{\mathrm{Z}}_n^*$ random variable expansion is:

$$
\begin{aligned}
&\sqrt{\frac{n_1 n_2}{n}}\left(\hat{\sigma}_h(\hat{\alpha},\hat{\beta})\hat{\beta}-\sigma_h\beta_0\right)\\
=&\sqrt{\frac{n_1 n_2}{n}}\left(\hat{\mu}_{h^2}(\alpha_0,\beta_0)-\mu_{h^2}\right)\frac{\beta_0}{\hat{\sigma}_h(\hat{\alpha},\hat{\beta})+\sigma_h}\\
-&\sqrt{\frac{n_1 n_2}{n}}\left(\hat{\mu}_h(\alpha_0,\beta_0)-\mu_h\right)\left(\frac{\hat{\mu}_h(\alpha_0,\beta_0)+\mu_h}{\hat{\sigma}_h(\hat{\alpha},\hat{\beta})+\sigma_h}\right)\beta_0\\
-&\sqrt{\frac{n_1 n_2}{n}}\frac{\mathbf{Q}_n}{\left(\hat{\sigma}_h(\hat{\alpha},\hat{\beta})+\sigma_h\right)}\left[\frac{1}{n}\,\boldsymbol{\nabla}\boldsymbol{\nabla}' l\left(\alpha,\beta\right)\Big|_{\left(\grave{\alpha},\grave{\beta}\right)}\right]^{-1}\frac{1}{n}\,\boldsymbol{\nabla} l\left(\alpha,\beta\right)\big|_{(\alpha_0,\beta_0)}
\end{aligned}
$$

which is written in vector notation as: $\tilde{\mathrm{Z}}_n^* = \boldsymbol{D}_n'\boldsymbol{Y}_n$

$$
\boldsymbol{D}_n \equiv \frac{1}{\hat{\sigma}_h\left(\hat{\alpha},\hat{\beta}\right)+\sigma_h}\begin{pmatrix} -\left(\hat{\mu}_h(\alpha_0,\beta_0)+\mu_h\right)\beta_0 \\ \beta_0 \\ -\left[\frac{1}{n}\,\boldsymbol{\nabla}\boldsymbol{\nabla}' l\left(\alpha,\beta\right)\big|_{\left(\grave{\alpha},\grave{\beta}\right)}\right]^{-1}\mathbf{Q}_n' \end{pmatrix} \tag{10}
$$

$$
\boldsymbol{Y}_n = \begin{pmatrix} Y_{1n} \\ Y_{2n} \\ Y_{3n} \\ Y_{4n} \end{pmatrix} \equiv \sqrt{\frac{n_1 n_2}{n}}\begin{pmatrix} \hat{\mu}_h\left(\alpha_0,\beta_0\right)-\mu_h \\ \hat{\mu}_{h^2}\left(\alpha_0,\beta_0\right)-\mu_{h^2} \\ \frac{1}{n}\,\boldsymbol{\nabla} l\left(\alpha,\beta\right)\big|_{(\alpha_0,\beta_0)} \end{pmatrix} \tag{11}
$$

where the gradient $\boldsymbol{\nabla} l(\alpha,\beta)\in\mathbb{R}^2$ is a column vector, the hessian $\boldsymbol{\nabla}\boldsymbol{\nabla}' l(\alpha,\beta)\in\mathbb{R}^{2\times 2}$ is a matrix, and $\mathbf{Q}_n\in\mathbb{R}^2$ is a row vector.

Assumption 2.1. The following list defines convergence conditions that allow $\tilde{\mathrm{Z}}_n^*$ to converge to a Gaussian random variable $\tilde{\mathrm{Z}}^*$:

- $h(x)$ is continuous and non-constant with respect to $g(x)$,
  i.e. $P_g(x : h(x) = m) = 0$ for all $m\in\mathbb{R}$.
- $h^k(x)$ is integrable with respect to $g_j(x)$ for $j = 1,\ldots,m$ and for $k = 1,2,3,4$.

The convergence conditions defined under Assumption 2.1 are used to show the following convergence results:

$$\left(\hat{\alpha},\hat{\beta}\right) \xrightarrow{P} (\alpha_0,\beta_0) \tag{12}$$

$$\hat{\mu}_h\left(\alpha_0,\beta_0\right) \xrightarrow{as} \mu_h \tag{13}$$

$$\hat{\sigma}_h^2\left(\hat{\alpha},\hat{\beta}\right) \xrightarrow{P} \sigma_h^2 \tag{14}$$

$$\boldsymbol{\nabla}\hat{\sigma}_h^2\left(\alpha,\beta\right)\Big|_{\left(\acute{\alpha},\acute{\beta}\right)} \xrightarrow{P} \boldsymbol{\nabla}\sigma_h^2\left(\alpha_0,\beta_0\right) \tag{15}$$

$$-\frac{1}{n}\,\boldsymbol{\nabla}\boldsymbol{\nabla}' l\left(\alpha,\beta\right)\Big|_{\left(\grave{\alpha},\grave{\beta}\right)} \xrightarrow{P} \mathbf{S}\left(\alpha_0,\beta_0\right) \tag{16}$$

$$\boldsymbol{Y}_n \xrightarrow{d} \boldsymbol{Y} \sim \mathrm{N}\left(\mathbf{0},\boldsymbol{\Sigma}\right). \tag{17}$$

The law of large numbers is applied in Lemma 2.1 and Corollary 2.1 to show the convergence result (13). The subsequent convergence results (14) through (16) are shown in Lemma 2.3 and Corollaries 2.4, 2.6, and 2.8 under the hypothesis that $(\hat{\alpha},\hat{\beta}),(\acute{\alpha},\acute{\beta}),(\grave{\alpha},\grave{\beta}) \xrightarrow{P} (\alpha_0,\beta_0)$. The convergence in probability result (12) is shown in Lemmas 2.4 though 2.6. The uniform convergence results of the abstract Glivenko-Cantelli theorem are applied in Lemmas 2.3 and 2.4 to show (12), (14), (15), and (16). The asymptotic properties of extremum estimators are applied in Lemma 2.4 to show (12). With regard to (15) and (16), the convergence in probability of $(\acute{\alpha},\acute{\beta})$ and $(\grave{\alpha},\grave{\beta})$ to (α_0,β_0) are shown in Corollary 2.9 as a consequence of $(\hat{\alpha},\hat{\beta})$ converging in probability to (α_0,β_0) from (12). The multivariate central limit theorem is applied in Lemma 2.8 to show (17). The convergence results, (12) through (16), are used together in Lemma 2.7 to show the limit in probability of $\boldsymbol{D}_n$. The convergence results for $\boldsymbol{D}_n$ and $\boldsymbol{Y}_n$ are used together in Theorem 2.2 to show the asymptotic distribution for $\tilde{\mathrm{Z}}_n^*$.

As described at the beginning of this section, the asymptotic distribution for $\tilde{\mathrm{Z}}_n^*$ is found for the case $(q=1, m=2)$. Note that some of the intermediate results, Lemmas 2.1 through 2.3, are shown for the general case $m = q+1 \geq 2$ since the extension is trivial. In Lemma 2.1, the law of large numbers is applied to show a generalization of (13).

Lemma 2.1. *For general $m > 1$, if a function $f(x)$ is integrable with respect to $g_j(x)$ for $j = 1,\ldots,m$, and if $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) = ((\alpha_{01},\ldots,\alpha_{0q})'),(\beta_{01},\ldots,\beta_{0q})')$ represents the true distortion parameters, then for $k = 1,\ldots,m$*

$$\sum_{i=1}^{n} f(t_i) w_k(t_i|\alpha_{0k},\beta_{0k})\hat{p}(t_i|\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0) \xrightarrow{as} Ef(X_k)\ . \tag{18}$$

Proof: The following weighted function of $f(x)$ for $k = 1, \ldots, m$ is integrable with respect to $g_j(x)$ for $j = 1 \ldots m$ since $f(x)$ is integrable by assumption

$$\left| \rho_j f(x) \frac{w_k(x|\alpha_{0k}, \beta_{0k})}{D_q(x|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)} \right| \leq \frac{\rho_j}{\rho_k} |f(x)|$$

Applying the law of large numbers, see van der Vaart (1998) [30] Example 2.1 and Proposition 2.16, shows that:

$$\begin{aligned}
&\sum_{i=1}^{n} f(t_i) w_k(t_i|\alpha_{0k}, \beta_{0k}) \hat{p}(t_i|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \\
&\qquad = \sum_{j=1}^{m} \frac{1}{n_j} \sum_{i=1}^{n_j} f(x_{ji}) w_k(x_{ji}|\alpha_{0k}, \beta_{0k}) \frac{\rho_j}{D_q(x_{ji}|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)} \\
&\qquad \overset{as}{\to} \sum_{j=1}^{m} \mathrm{E}\left(f(X_j) w_k(X_j|\alpha_{0k}, \beta_{0k}) \frac{\rho_j}{D_q(X_j|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)} \right) \\
&\qquad = \mathrm{E}\left(f(X_m) w_k(X_m|\alpha_{0k}, \beta_{0k}) \right) \\
&\qquad = \mathrm{E} f(X_k)\ . \ \blacksquare
\end{aligned}$$

Corollary 2.1. *For $m = 2$ with $k = m$, if $h(x)$ is integrable with respect to $g_j(x)$ for $j = 1, 2$, and if (α_0, β_0) represents the true distortion parameters, then $\hat{\mu}_h(\alpha_0, \beta_0) \overset{as}{\to} \mu_h$, proving (13).* $\blacksquare$

The abstract Glivenko-Cantelli theorem is applied to establish uniform convergence results for a class of parametric functions. The following Definitions 2.1 and 2.2, Theorem 2.1, and Example 2.1, are taken from van der Vaart (1998) [30] section 19.2.

Definition 2.1. A class $\mathcal{F}$ of measurable integrable functions f is called P-Glivenko-Cantelli if

$$\|\mathbb{P}_n f - Pf\|_{\mathcal{F}} \equiv \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(x_i) - \int f dP \right| \overset{as*}{\to} 0$$

or equivalently, if there exists a sequence of random variables $\triangle_n$ such that

$$\|\mathbb{P}_n f - Pf\|_{\mathcal{F}} \leq \triangle_n \text{ and } \triangle_n \overset{as}{\to} 0$$

where $x_1, \ldots, x_n$ is a random sample from the probability distribution P.

Definition 2.2. Given two functions l and u, the bracket$[l, u]$ is the set of all functions f with $l \leq f \leq u$. An ε-bracket in $L_r(P)$ is a bracket$[l, u]$ with $P(u-l)^r < \varepsilon^r$. The bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, L_r(P))$ is the minimum number of ε-brackets needed to cover $\mathcal{F}$. The bracketing functions l and u must have finite $L_r(P)$-norms but need not belong to $\mathcal{F}$.

Theorem 2.1 (Abstract Glivenko-Cantelli). *Every class $\mathcal{F}$ of measurable [integrable] functions such that $N_{[\,]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\varepsilon > 0$ is P-Glivenko-Cantelli.* ■

Example 2.1 (Parametric Class). Let $\mathcal{F} = \{f_{\boldsymbol{\theta}} \in L_1(P) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ be a collection of measurable [integrable] functions indexed by a bounded subset $\boldsymbol{\Theta} \subset \mathbb{R}^d$. Suppose that there exists a measurable function m such that

$$|f_{\boldsymbol{\theta}_1}(x) - f_{\boldsymbol{\theta}_2}(x)| \leq m(x)\, \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|, \text{ every } \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}\,.$$

If $\|m\|_{P,r}^r \equiv P|m|^r < \infty$, then there exists a constant K, depending on $\boldsymbol{\Theta}$ and d only, such that the bracketing numbers satisfy

$$N_{[\,]}(\varepsilon\|m\|_{P,r}, \mathcal{F}, L_r(P)) \leq K \left(\frac{\text{diam } \boldsymbol{\Theta}}{\varepsilon} \right)^d, \text{ every } 0 < \varepsilon < \text{ diam } \boldsymbol{\Theta}\,.$$

The Lipschitz condition shows that $f_{\boldsymbol{\theta}_1} - \varepsilon m \leq f_{\boldsymbol{\theta}_2} \leq f_{\boldsymbol{\theta}_1} + \varepsilon m$ if $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \leq \varepsilon$. Hence a $2\varepsilon\|m\|_{P,r}$-bracket in $L_r(P)$ for the parametric class of functions $\mathcal{F}$ takes the form $[f_{\boldsymbol{\theta}} - \varepsilon m, f_{\boldsymbol{\theta}} + \varepsilon m]$. ■

Thus the bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, L_1(P))$ in Example 2.1 is finite for every $\varepsilon > 0$ and the class of integrable functions $\mathcal{F}$ is P-Glivenko-Cantelli.

The abstract Glivenko-Cantelli Theorem 2.1 for a parametric class from Example 2.1, is applied to establish uniform convergence results as defined by (20) below, for a class of integrable functions parameterized by $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, when an integrable Lipschitz condition is met as defined by (19) below.

Lemma 2.2. *For general $m > 1$, let $\mathcal{F}_j \equiv \{f(\cdot|\boldsymbol{\alpha}, \boldsymbol{\beta}) \in L_1(G_j) : (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\Theta}\}$ for $j = 1, \ldots, m$ denote m parametric classes of functions where each class denotes a collection of functions indexed by a bounded subset $\boldsymbol{\Theta} \subset \mathbb{R}^{2q}$ that are integrable with respect to the probability distributions G_j associated with the densities g_j. If $f(\cdot|\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{F}_j$ has an integrable Lipschitz bound $m_j(\cdot)$ with respect to G_j as defined by*

$$\begin{aligned} &\left| f(x|\boldsymbol{\alpha}^1, \boldsymbol{\beta}^1) - f(x|\boldsymbol{\alpha}^2, \boldsymbol{\beta}^2) \right| \leq m_j(x) \left\| (\boldsymbol{\alpha}^1, \boldsymbol{\beta}^1)' - (\boldsymbol{\alpha}^2, \boldsymbol{\beta}^2)' \right\| \\ &\qquad \textit{for every } (\boldsymbol{\alpha}^1, \boldsymbol{\beta}^1), (\boldsymbol{\alpha}^2, \boldsymbol{\beta}^2) \in \boldsymbol{\Theta} \\ &\qquad E(m_j(X_j)) < \infty \textit{ for } j \in \{1 \ldots m\} \end{aligned} \tag{19}$$

then each class $\mathcal{F}_j$ is G_j-Glivenko-Cantelli, by the abstract Glivenko-Cantelli Theorem 2.1 as applied in Example 2.1 to a parametric class of functions, resulting in uniform convergence almost surely for all functions $f \in \mathcal{F}_j$

$$\|\mathbb{P}_{n_j} f - Pf\|_{\mathcal{F}_j} \equiv \sup_{f \in \mathcal{F}_j} \left| \frac{1}{n_j} \sum_{i=1}^{n_j} f(x_{ji}|\boldsymbol{\alpha}, \boldsymbol{\beta}) - E(f(X_j|\boldsymbol{\alpha}, \boldsymbol{\beta})) \right| \xrightarrow{as*} 0. \blacksquare \quad (20)$$

Definition 2.3. For general $m > 1$, let $\mathcal{F}_j(f_1, f_2)$ for $j = 1, \ldots, m$ denote m parametric classes of functions as defined below that are indexed by a bounded subset $\boldsymbol{\Theta} \subset \mathbb{R}^{2q}$ which contains the true distortion parameters $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and that are integrable with respect to the probability distributions G_j associated with the densities g_j

$$\mathcal{F}_j(f_1, f_2) \equiv \{f(\cdot|\boldsymbol{\alpha}, \boldsymbol{\beta}) = f_1(\cdot) f_2(\cdot|\boldsymbol{\alpha}, \boldsymbol{\beta}) \frac{\rho_j}{D_q(\cdot|\boldsymbol{\alpha}, \boldsymbol{\beta})} : (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\Theta}\} \quad (21)$$

where $f_1 \in L_1(G_j)$, $f_2 \in L_\infty(G_j)$, and $f \in \mathcal{F}_j \subset L_1(G_j)$.

Definition 2.3 associates m abstract parametric classes of integrable functions with each of the m densities $\{g_1, \ldots, g_m\}$. This structure allows the abstract Glivenko-Cantelli theorem to be applied to a random sample from each of the densities in order to show a uniform law of large numbers convergence result over the functions in each class. At this time the function parameters of each class, f_1 and f_2, have only been defined in the abstract. Each of these function parameters are specialized in Definitions 2.4 and 2.5 to well defined functions in order to show specific uniform law of large numbers convergence results. The parametric index $\boldsymbol{\Theta}$ describes any bounded subset of $\mathbb{R}^{2q}$ such that each resulting class of indexed functions $\mathcal{F}_j(f_1, f_2)$ for $j = 1, \ldots, m$ meets the integrable conditions imposed on f_1, f_2, and f. In the subsequent analysis, the parametric index $\boldsymbol{\Theta}$ will be specialized as needed to show each of the convergence results (12), (14), (15), and (16).

Corollary 2.2. *Under the conditions of Lemma 2.2 with $\mathcal{F}_j$ specialized to $\mathcal{F}_j(f_1, f_2)$ with parametric index $\boldsymbol{\Theta}$ from Definition 2.3, applying (20) or applying the law of large numbers, for any fixed $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \boldsymbol{\Theta}$, shows*

$$\begin{aligned} &\sum_{i=1}^{n} f_1(t_i) f_2(t_i|\boldsymbol{\alpha}, \boldsymbol{\beta}) \hat{p}(t_i|\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &= \sum_{j=1}^{m} \frac{1}{n_j} \sum_{i=1}^{n_j} f_1(x_{ji}) f_2(x_{ji}|\boldsymbol{\alpha}, \boldsymbol{\beta}) \frac{\rho_j}{D_q(x_{ji}|\boldsymbol{\alpha}, \boldsymbol{\beta})} \\ &\xrightarrow{as} \sum_{j=1}^{m} E\left(f_1(X_j) f_2(X_j|\boldsymbol{\alpha}, \boldsymbol{\beta}) \frac{\rho_j}{D_q(X_j|\boldsymbol{\alpha}, \boldsymbol{\beta})} \right). \blacksquare \end{aligned} \quad (22)$$

Lemma 2.3. *Under the conditions of Lemma 2.2 with $\mathcal{F}_j$ specialized to $\mathcal{F}_j(f_1, f_2)$ with parametric index $\boldsymbol{\Theta}$ from Definition 2.3, if $(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*) \xrightarrow{P} (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) \in \boldsymbol{\Theta}$ then*

$$\sum_{i=1}^{n} f_1\left(t_i\right) f_2\left(t_i|\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*\right) \hat{p}\left(t_i|\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*\right) \xrightarrow{P} E\left(f_1\left(X_m\right) f_2\left(X_m|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\right)\right) .$$

Proof: For any random sequence $(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*) \xrightarrow{P} (\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1) \in \boldsymbol{\Theta}$ as $n \to \infty$, applying (20) from Lemma 2.2 or the law of large numbers for $(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)$, and applying Slutsky's theorem shows

$$\begin{aligned}
&\left|\frac{1}{n_j}\sum_{i=1}^{n_j} f\left(x_{ji}|\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*\right) - \mathrm{E}\left(f\left(X_j|\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1\right)\right)\right| \\
\leq &\left|\frac{1}{n_j}\sum_{i=1}^{n_j} f\left(x_{ji}|\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1\right) - \mathrm{E}\left(f\left(X_j|\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1\right)\right)\right| \\
+ &\frac{1}{n_j}\sum_{i=1}^{n_j} m_j\left(x_{ji}\right) \left\|(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*)' - (\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)'\right\| \\
\xrightarrow{P} &\, 0 .
\end{aligned} \tag{23}$$

Consequently, as $(\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*) \xrightarrow{P} (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, the general convergence in probability result follows, that

$$\begin{aligned}
&\sum_{i=1}^{n} f_1\left(t_i\right) f_2\left(t_i|\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*\right) \hat{p}\left(t_i|\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*\right) \\
= &\sum_{j=1}^{m} \frac{1}{n_j}\sum_{i=1}^{n_j} f_1\left(x_{ji}\right) f_2\left(x_{ji}|\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*\right) \frac{\rho_j}{D_q\left(x_{ji}|\boldsymbol{\alpha}_*, \boldsymbol{\beta}_*\right)} \\
\xrightarrow{P} &\sum_{j=1}^{m} \mathrm{E}\left(f_1(X_j) f_2(X_j|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\frac{\rho_j}{D_q(X_j|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)}\right) \\
= &\,\mathrm{E}\left(f_1(X_m) f_2(X_m|\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\right) . \blacksquare
\end{aligned} \tag{24}$$

Definition 2.4. For $m = 2$, let $\mathcal{F}_{j|k}^{(1)}(\boldsymbol{\Theta}) \equiv \mathcal{F}_j(h^k(x), 1)$ with parametric index $\boldsymbol{\Theta} \subset \mathbb{R}^2$ for $k = 0, 1, 2$ and $j = 1, 2$ define 6 classes of integrable functions that are specialized versions of $\mathcal{F}_j(f_1, f_2)$ from Definition 2.3.

Remark 2.1. The function $f(x|\alpha, \beta) \in \mathcal{F}_{j|k}^{(1)}(\boldsymbol{\Theta})$ has partial derivatives of all orders with respect to (α, β). A Taylor series expansion for $f(x|\alpha, \beta) \in$

$\mathcal{F}^{(1)}_{j|k}(\mathbf{\Theta})$ around $(\alpha, \beta) \in \mathbf{\Theta}$ given the gradient $\mathbf{\nabla} \equiv (\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta})'$, and the mean value theorem 6.7 [16], are used to find a Lipschitz bound that depends on (α, β) and on the maximum vector norm $\|v\|_\infty \equiv \max_i |v_i|$

$$\begin{aligned}
f\left(x|\alpha^1, \beta^1\right) - f\left(x|\alpha^2, \beta^2\right) &= \mathbf{\nabla}' f\left(x|\alpha^*, \beta^*\right)\left[\left(\alpha^1, \beta^1\right)' - \left(\alpha^2, \beta^2\right)'\right] \\
\left|f\left(x|\alpha^1, \beta^1\right) - f\left(x|\alpha^2, \beta^2\right)\right| &\leq \max_{1 \leq \lambda \leq 1} \left\|\mathbf{\nabla}' f\left(x|\alpha_\lambda, \beta_\lambda\right)\right\|_\infty \left\|\left(\alpha^1, \beta^1\right)' - \left(\alpha^2, \beta^2\right)'\right\|_\infty \\
\left(\alpha_\lambda, \beta_\lambda\right) &= \lambda\left(\alpha^1, \beta^1\right) + (1-\lambda)\left(\alpha^2, \beta^2\right), \ \lambda \in [0,1] \\
\left(\alpha^*, \beta^*\right) &= \left(\alpha_{\lambda^*}, \beta_{\lambda^*}\right) \ \text{for some } \lambda^* \in (0,1) \ .
\end{aligned}$$

The previous display leads to an integrable Lipschitz bound $m^{(1)}_{j|k}(x)$ that does not depend on (α, β)

$$\begin{aligned}
\forall \ (\alpha, \beta) \in \mathbf{\Theta}: \ \left\|\mathbf{\nabla}' f\left(x|\alpha, \beta\right)\right\|_\infty &= \left\|-\rho_j h^k\left(x\right) \frac{\rho_1 w_1\left(x|\alpha, \beta\right)}{D_1^2\left(x|\alpha, \beta\right)}\left(1, h\left(x\right)\right)\right\|_\infty \\
&\leq \rho_j \left|h^k\left(x\right)\right| \left\|\left(1, h\left(x\right)\right)\right\|_\infty \\
&\leq \rho_j \left(\left|h^k\left(x\right)\right| + \left|h^{k+1}\left(x\right)\right|\right) \\
&\equiv m^{(1)}_{j|k}(x) \ . \qquad (25)
\end{aligned}$$

Given any bounded subset $\mathbf{\Theta} \subset \mathbb{R}^2$, it is easy to show that the integrable conditions of Definition 2.3 are met since for any $f(x|\alpha, \beta) \in \mathcal{F}^{(1)}_{j|k}(\mathbf{\Theta})$ with $(\alpha, \beta) \in \mathbf{\Theta}$ and with $j = 1, 2$ and $k = 0, 1, 2$

$$\left|f\left(x|\alpha, \beta\right)\right| \leq \left|f\left(x|0, 0\right)\right| + m^{(1)}_{j|k}\left(x\right)\left\|\left(\alpha, \beta\right)\right\|_\infty \ .$$

Hence $f_1(x) \equiv h^k(x) \in L_1(G_j)$, $f(x|\alpha, \beta) \in L_1(G_j)$, and $m^{(1)}_{j|k}(x) \in L_1(G_j)$ for $j = 1, 2$ and $k = 0, 1, 2$ under the convergence conditions of Assumption 2.1. Also $f_2(x|\alpha, \beta) \equiv 1 \in L_\infty(G_j)$ for $j = 1, 2$.

Corollary 2.3. *Under the conditions of Lemma 2.3 with $\mathcal{F}_j(f_1, f_2)$ specialized to $\mathcal{F}^{(1)}_{j|k}(\mathbf{\Theta}) \equiv \mathcal{F}_j(h^k(x), 1)$ with parametric index $\mathbf{\Theta}$ from Definition 2.4 with $j = 1, 2$ and $k = 0, 1, 2$, if $h^l(x)$ is integrable with respect to $g_j(x)$ for $j = 1, 2$ and $l = 0, 1, 2, 3$, then for any fixed $(\alpha, \beta) \in \mathbf{\Theta}$ and for any sequence $(\alpha_*, \beta_*) \xrightarrow{P} (\alpha_0, \beta_0) \in \mathbf{\Theta}$*

$$\hat{\mu}_{h^k}\left(\alpha, \beta\right) = \sum_{i=1}^n h^k(t_i)\hat{p}(t_i|\alpha, \beta) \xrightarrow{as} \sum_{j=1}^2 E\left(\frac{\rho_j h^k\left(X_j\right)}{D_1\left(X_j|\alpha, \beta\right)}\right) \qquad (26)$$

$$\hat{\mu}_{h^k}\left(\alpha_*, \beta_*\right) \xrightarrow{P} \mu_{h^k} \ . \qquad (27)$$

Proof: Under the assumptions, $f(x|\alpha,\beta) \in \mathcal{F}^{(1)}_{j|k}(\mathbf{\Theta})$ is integrable with respect to $g_j(x)$ for $j = 1, 2$ and $k = 0, 1, 2$, and $m^{(1)}_{j|k}(x)$ is integrable with respect to $g_j(x)$ for $j = 1, 2$ and $k = 0, 1, 2$ so that the integrable Lipschitz condition (19) is met. Hence the results of Corollary 2.2 are valid for any fixed $(\alpha, \beta) \in \mathbf{\Theta}$ and the results of Lemma 2.3 are valid for any sequence $(\alpha_*, \beta_*) \xrightarrow{P} (\alpha_0, \beta_0) \in \mathbf{\Theta}$. ■

Corollary 2.4. *Under the conditions of Corollary 2.3, if* $(\hat{\alpha}, \hat{\beta}) \xrightarrow{P} (\alpha_0, \beta_0) \in \mathbf{\Theta}$*, then* $\hat{\mu}_{h^k}(\hat{\alpha}, \hat{\beta}) \xrightarrow{P} \mu_{h^k}$ *for* $k = 1, 2$*. Hence* $\hat{\sigma}^2_h(\hat{\alpha}, \hat{\beta}) \xrightarrow{P} \sigma^2_h$*, proving (14).* ■

To analyze $\mathbf{Q}_n((\hat{\alpha}, \hat{\beta}), (\acute{\alpha}, \acute{\beta}))$, previously defined in (7), as $(\hat{\alpha}, \hat{\beta}), (\acute{\alpha}, \acute{\beta}) \xrightarrow{P} (\alpha_0, \beta_0)$, the convergence in probability of $\nabla\hat{\sigma}^2_h(\alpha, \beta)\big|_{(\acute{\alpha}, \acute{\beta})}$ is shown. Note that the convergence in probability of $\hat{\sigma}^2_h(\hat{\alpha}, \hat{\beta})$, has already been proven in the previous Corollary 2.4.

With regard to convergence in probability of $\nabla\hat{\sigma}^2_h(\alpha, \beta)\big|_{(\acute{\alpha}, \acute{\beta})}$, the definition of $\hat{\sigma}^2_h(\alpha, \beta)$ is used to find $\nabla\hat{\sigma}^2_h(\alpha, \beta)$ as follows

$$\hat{\sigma}^2_h(\alpha, \beta) = \sum_{i=1}^{n} h^2(t_i)\hat{p}(t_i|\alpha, \beta) - \left(\sum_{i=1}^{n} h(t_i)\hat{p}(t_i|\alpha, \beta)\right)^2 \tag{28}$$

$$\begin{aligned} \frac{\partial}{\partial\alpha}\hat{\sigma}^2_h(\alpha, \beta) = &-\sum_{i=1}^{n} h^2(t_i)\hat{p}^2(t_i|\alpha, \beta)w_1(t_i|\alpha, \beta)n_1 \\ &+ 2\hat{\mu}_h(\alpha, \beta)\left(\sum_{i=1}^{n} h(t_i)\hat{p}^2(t_i|\alpha, \beta)w_1(t_i|\alpha, \beta)n_1\right) \end{aligned} \tag{29}$$

$$\begin{aligned} \frac{\partial}{\partial\beta}\hat{\sigma}^2_h(\alpha, \beta) = &-\sum_{i=1}^{n} h^3(t_i)\hat{p}^2(t_i|\alpha, \beta)w_1(t_i|\alpha, \beta)n_1 \\ &+ 2\hat{\mu}_h(\alpha, \beta)\left(\sum_{i=1}^{n} h^2(t_i)\hat{p}^2(t_i|\alpha, \beta)w_1(t_i|\alpha, \beta)n_1\right). \end{aligned} \tag{30}$$

Definition 2.5. For $m = 2$, let $\mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta}) \equiv \mathcal{F}_j(h^k(x), \rho_1 w_1(x|\alpha, \beta)/D_1(x|\alpha, \beta))$ with parametric index $\mathbf{\Theta} \subset \mathbb{R}^2$ for $k = 0, 1, 2, 3$ and $j = 1, 2$ define 8 classes of integrable functions that are specialized versions of $\mathcal{F}_j(f_1, f_2)$ from Definition 2.3.

Remark 2.2. The function $f(x|\alpha, \beta) \in \mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta})$ has partial derivatives of all orders with respect to (α, β). A Lipschitz bound $m^{(2)}_{j|k}(x)$ is found, by using a Taylor series expansion for $f(x|\alpha,\beta) \in \mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta})$ around $(\alpha,\beta) \in \mathbf{\Theta}$ given the gradient $\mathbf{\nabla} \equiv (\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta})'$, by using the mean value theorem 6.7 [16], and by using the maximum vector norm $\|\cdot\|_\infty$

$$
\begin{aligned}
\forall\ (\alpha,\beta) \in \mathbf{\Theta}:\ & \left\|\mathbf{\nabla}' f\left(x|\alpha,\beta\right)\right\|_\infty \\
&= \left\|\rho_j h^k\left(x\right) \frac{\rho_1 w_1\left(x|\alpha,\beta\right)}{D_1^2\left(x|\alpha,\beta\right)} \left(1 - 2\frac{\rho_1 w_1\left(x|\alpha,\beta\right)}{D_1\left(x|\alpha,\beta\right)}\right)\left(1, h\left(x\right)\right)\right\|_\infty \\
&\leq \rho_j \left|h^k\left(x\right)\right| (1)\,(3)\,\left\|(1, h\left(x\right))\right\|_\infty \\
&\leq 3\rho_j \left(\left|h^k\left(x\right)\right| + \left|h^{k+1}\left(x\right)\right|\right) \\
&\equiv m^{(2)}_{j|k}\left(x\right)\ . \qquad (31)
\end{aligned}
$$

Given any bounded subset $\mathbf{\Theta} \subset \mathbb{R}^2$, it is easy to show that the integrable conditions of Definition 2.3 are met since for any $f(x|\alpha,\beta) \in \mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta})$ with $(\alpha,\beta) \in \mathbf{\Theta}$ and with $j = 1,2$ and $k = 0,1,2$

$$
\left|f\left(x|\alpha,\beta\right)\right| \leq \left|f\left(x|0,0\right)\right| + m^{(2)}_{j|k}\left(x\right)\left\|(\alpha,\beta)\right\|_\infty\ .
$$

Hence $f_1(x) \equiv h^k(x) \in L_1(G_j)$, $f(x|\alpha,\beta) \in L_1(G_j)$, and $m^{(2)}_{j|k}(x) \in L_1(G_j)$ for $j = 1,2$ and $k = 0,1,2$ under the convergence conditions of Assumption 2.1. Also $f_2(x|\alpha,\beta) \equiv \rho_1 w_1(x|\alpha,\beta)/D_1(x|\alpha,\beta) \in L_\infty(G_j)$ for $j = 1,2$.

Corollary 2.5. *Under the conditions of Lemma 2.3 with $\mathcal{F}_j(f_1,f_2)$ specialized to $\mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta}) \equiv \mathcal{F}_j(h^k(x), \rho_1 w_1(x|\alpha,\beta)/D_1(x|\alpha,\beta))$ with parametric index $\mathbf{\Theta}$ from Definition 2.5 with $j = 1,2$ and $k = 0,1,2,3$, if $h^l(x)$ is integrable with respect to $g_j(x)$ for $j = 1,2$ and $l = 0,1,2,3,4$, and $(\alpha_*,\beta_*) \xrightarrow{P} (\alpha_0,\beta_0) \in \mathbf{\Theta}$ then*

$$
\begin{aligned}
\sum_{i=1}^{n} h^k(t_i)\hat{p}^2(t_i|\alpha_*,\beta_*)w_1(t_i|\alpha_*,\beta_*)n_1 &\xrightarrow{P} \rho_1 E\left(\frac{h^k(X_2)w_1\left(X_2|\alpha_0,\beta_0\right)}{D_1\left(X_2|\alpha_0,\beta_0\right)}\right) \\
&= \rho_1 E\left(\frac{h^k(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right)\ .
\end{aligned}
$$

Proof: Under the assumptions, $f(x|\alpha,\beta) \in \mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta})$ is integrable with respect to $g_j(x)$ for $j = 1,2$ and $k = 0,1,2,3$, and $m^{(2)}_{j|k}(x)$ is integrable with respect to $g_j(x)$ for $j = 1,2$ and $k = 0,1,2,3$ so that the integrable Lipschitz condition (19) is met. Hence the results of Lemma 2.3 are valid for $(\alpha_*,\beta_*) \xrightarrow{P} (\alpha_0,\beta_0) \in \mathbf{\Theta}$. ■

Corollary 2.6. *Under the conditions of Corollary 2.5, if* $(\acute{\alpha},\acute{\beta}) \xrightarrow{P} (\alpha_0,\beta_0) \in \mathbf{\Theta}$*, then*

$$\begin{aligned}\nabla\hat{\sigma}_h^2(\alpha,\beta)\big|_{(\acute{\alpha},\acute{\beta})} &\xrightarrow{P} \rho_1 \begin{pmatrix} 2\mu_h E\left(\frac{h(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) - E\left(\frac{h^2(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) \\ 2\mu_h E\left(\frac{h^2(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) - E\left(\frac{h^3(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right)\end{pmatrix} \\ &\equiv \nabla\sigma_h^2(\alpha_0,\beta_0)\end{aligned}$$

proving (15). ■

Corollary 2.7. *Under the conditions of Corollaries 2.4 and 2.6, applying (14) and (15) shows that* $\mathbf{Q}_n((\hat{\alpha},\hat{\beta}),(\acute{\alpha},\acute{\beta}))$ *from (7) converges in probability to* $\mathbf{Q}(\alpha_0,\beta_0)$ *defined as*

$$\mathbf{Q}(\alpha_0,\beta_0) = \begin{pmatrix} \rho_1\beta_0\left[2\mu_h E\left(\frac{h(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) - E\left(\frac{h^2(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right)\right] \\ 2\sigma_h^2 + \rho_1\beta_0\left[2\mu_h E\left(\frac{h^2(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) - E\left(\frac{h^3(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right)\right]\end{pmatrix}' . \blacksquare$$

The convergence in probability of $-n^{-1}\nabla\nabla' l(\alpha,\beta)\big|_{(\grave{\alpha},\grave{\beta})}$ to $\mathbf{S}(\alpha_0,\beta_0)$ is shown by using the almost sure convergence of functions in the previously defined classes of functions $f \in \mathcal{F}_{j|k}^{(2)}(\mathbf{\Theta}) \equiv \mathcal{F}_j(h^k(x), \rho_1 w_1(x|\alpha,\beta)/D_1(x|\alpha,\beta))$ with parametric index $\mathbf{\Theta}$ from Definition 2.5 where $j = 1,2$ and $k = 1,2$.

$$\begin{aligned}\frac{1}{n}\nabla l(\alpha,\beta) &= \frac{\rho_1}{1+\rho_1}\begin{pmatrix} 1 - \sum_{i=1}^n \hat{p}(t_i|\alpha,\beta)w_1(t_i|\alpha,\beta) \\ \frac{1}{n_1}\sum_{i=1}^{n_1} h(x_{1i}) - \sum_{i=1}^n h(t_i)\hat{p}(t_i|\alpha,\beta)w_1(t_i|\alpha,\beta)\end{pmatrix} \\ &= \frac{1}{1+\rho_1}\begin{pmatrix} \sum_{i=1}^n \hat{p}(t_i|\alpha,\beta) - 1 \\ \sum_{i=1}^n h(t_i)\hat{p}(t_i|\alpha,\beta) - \frac{1}{n_2}\sum_{i=1}^{n_2} h(x_{2i})\end{pmatrix}\end{aligned} \tag{32}$$

The components of $\nabla\nabla' l(\alpha,\beta)/n$ are

$$\begin{aligned}\frac{\partial^2}{\partial\alpha^2}\frac{l(\alpha,\beta)}{n} &= -\frac{1}{1+\rho_1}\left(\sum_{i=1}^n \hat{p}^2(t_i|\alpha,\beta)w_1(t_i|\alpha,\beta)n_1\right) \\ \frac{\partial^2}{\partial\alpha\partial\beta}\frac{l(\alpha,\beta)}{n} &= -\frac{1}{1+\rho_1}\left(\sum_{i=1}^n h(t_i)\hat{p}^2(t_i|\alpha,\beta)w_1(t_i|\alpha,\beta)n_1\right) \\ \frac{\partial^2}{\partial\beta^2}\frac{l(\alpha,\beta)}{n} &= -\frac{1}{1+\rho_1}\left(\sum_{i=1}^n h^2(t_i)\hat{p}^2(t_i|\alpha,\beta)w_1(t_i|\alpha,\beta)n_1\right)\end{aligned} \tag{33}$$

Corollary 2.8. *Under the conditions of Corollary 2.5, if* $(\grave{\alpha}, \grave{\beta}) \xrightarrow{P} (\alpha_0, \beta_0) \in \mathbf{\Theta}$*, then*

$$-\frac{1}{n}\,\boldsymbol{\nabla}\boldsymbol{\nabla}' l\,(\alpha,\beta)\big|_{(\alpha_*,\beta_*)} \xrightarrow{P} \frac{\rho_1}{1+\rho_1}\begin{bmatrix} E\left(\frac{1}{D_1(X_1|\alpha_0,\beta_0)}\right) & E\left(\frac{h(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) \\ E\left(\frac{h(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) & E\left(\frac{h^2(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right)\end{bmatrix} \tag{34}$$

$$= \mathbf{S}\,(\alpha_0,\beta_0)$$

$$-\frac{1}{n}\,\boldsymbol{\nabla}\boldsymbol{\nabla}' l\,(\alpha,\beta)\big|_{(\grave{\alpha},\grave{\beta})} \xrightarrow{P} \mathbf{S}\,(\alpha_0,\beta_0) \tag{35}$$

The previous display (35) proves (16). ■

To complete the convergence in probability analysis of $\boldsymbol{D}_n$, the convergence in probability of $(\hat{\alpha}, \hat{\beta})$ to (α_0, β_0) is shown using the asymptotic properties of extremum estimators as developed by Amemiya (1985) [1]. Definition 4.1.1, in Amemiya [1], defines three modes of uniform convergence to $\mathbf{0}$ for a non-negative sequence of random variables $g_T(\boldsymbol{\theta})$ that depend on a parameter vector $\boldsymbol{\theta}$.

(i) $P(\lim_{T\to\infty} \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} g_T(\boldsymbol{\theta}) = \mathbf{0}) = 1$ is described as convergence almost surely uniformly in $\boldsymbol{\theta} \in \mathbf{\Theta}$.

(ii) $\lim_{T\to\infty} P(\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} g_T(\boldsymbol{\theta}) < \boldsymbol{\epsilon}) = 1$ for any $\boldsymbol{\epsilon} > \mathbf{0}$ is described as convergence in probability uniformly in $\boldsymbol{\theta} \in \mathbf{\Theta}$.

(iii) $\lim_{T\to\infty} \inf_{\boldsymbol{\theta}\in\mathbf{\Theta}} P(g_T(\boldsymbol{\theta}) < \boldsymbol{\epsilon}) = 1$ for any $\boldsymbol{\epsilon} > \mathbf{0}$ is described as convergence in probability semiuniformly in $\boldsymbol{\theta} \in \mathbf{\Theta}$.

As reported in Amemiya [1], the first mode of uniform convergence (i) implies the second mode (ii) and the second mode (ii) implies the third mode (iii). The first mode of uniform convergence (i), is equivalent to the almost sure convergence of the functions, $f \in \mathcal{F}_j$ for $j = 1 \ldots m$, as shown in (20). The second mode of uniform convergence (ii), is one condition of Theorem 4.1.6 (out of six conditions), in Amemiya [1], to show that an extremum estimator converges in probability to the actual parameter.

In order to apply the theory of extremum estimators, the stochastic function $l_n(\alpha,\beta) = l(\alpha,\beta) + n\log(n_2)$ is identified with $g_T(\boldsymbol{\theta})$, where maximizing $l_n(\alpha,\beta)$ with respect to (α,β) is equivalent to maximizing $l(\alpha,\beta)$ with respect to (α,β), since the difference between $l_n(\alpha,\beta)$ and $l(\alpha,\beta)$, $n\log(n_2)$, is a constant relative to (α,β). Let $\mathbf{\Theta}_n = \{(\alpha_*,\beta_*) : \boldsymbol{\nabla} l_n(\alpha_*,\beta_*) = \mathbf{0}\}$ so that $(\hat{\alpha},\hat{\beta}) \in \mathbf{\Theta}_n$.

Lemma 2.4. *If $h^k(x)$ is integrable with respect to $g_j(x)$ for $j = 1, 2$ and $k = 1, 2$ and if $h(x)$ is non-constant with respect to $g_2(x)$ then one of the roots $(\alpha_*, \beta_*) \in \mathbf{\Theta}_n$ converges in probability to (α_0, β_0).*

Proof: Let $\mathbf{\Theta}$ denote an open bounded convex subset of $\mathbb{R}^2$ containing (α_0, β_0). Application of Theorem 4.1.6, from Amemiya [1], shows that the result is true under the following conditions:

(A) $\boldsymbol{\nabla\nabla}' l_n(\alpha, \beta)$ exists and is continuous for $(\alpha, \beta) \in \mathbf{\Theta}$ an open convex neighborhood of (α_0, β_0),

(B) $n^{-1}\boldsymbol{\nabla\nabla}' l_n(\alpha, \beta)|_{(\alpha_*, \beta_*)}$ converges in probability to a finite nonsingular matrix $-\mathbf{S}(\alpha_0, \beta_0) = \lim\ n^{-1}\mathrm{E}\boldsymbol{\nabla\nabla}' l_n(\alpha, \beta)|_{(\alpha_0, \beta_0)}$ for any sequence (α_*, β_*) converging in probability to (α_0, β_0),

(C) $n^{-1/2}\boldsymbol{\nabla} l_n(\alpha, \beta)|_{(\alpha_0, \beta_0)} \to \mathrm{N}(\mathbf{0}, \boldsymbol{B}(\alpha_0, \beta_0))$
where $\boldsymbol{B}(\alpha_0, \beta_0) = \lim\ n^{-1}\mathrm{E}(\boldsymbol{\nabla} l_n(\alpha, \beta)|_{(\alpha_0, \beta_0)}) \times (\boldsymbol{\nabla}' l_n(\alpha, \beta)|_{(\alpha_0, \beta_0)})$,

(D) $n^{-1} l_n(\alpha, \beta)$ converges to a nonstochastic function in probability uniformly in $(\alpha, \beta) \in \mathbf{\Theta}$ an open neighborhood of (α_0, β_0),

(E) $-\mathbf{S}(\alpha_0, \beta_0)$ defined in condition (B) is a negative definite matrix,

(F) The limit in probability of $n^{-1}\boldsymbol{\nabla\nabla}' l_n(\alpha, \beta)$ exists and is continuous for $(\alpha, \beta) \in \mathbf{\Theta}$ a neighborhood of (α_0, β_0).

Condition (A) is immediate after examining (33). Condition (B) is proven by starting with a consequence (34) from Corollary 2.8 of the abstract Glivenko-Cantelli Theorem 2.1 for a parametric class with a parametric index $\mathbf{\Theta}$ and by applying a result of the law of large numbers (18) from Lemma 2.1, in order to show

$$\begin{aligned}
&\frac{1}{n}\ \boldsymbol{\nabla\nabla}' l_n\left(\alpha, \beta\right)\Big|_{(\alpha_*, \beta_*)} \xrightarrow{P} -\mathbf{S}\left(\alpha_0, \beta_0\right) \ \text{ as } \ (\alpha_*, \beta_*) \xrightarrow{P} (\alpha_0, \beta_0) \\
&\frac{1}{n}\ \boldsymbol{\nabla\nabla}' l_n\left(\alpha, \beta\right)\Big|_{(\alpha_0, \beta_0)} \xrightarrow{as} -\mathbf{S}(\alpha_0, \beta_0)
\end{aligned}$$

and by direct calculation to show

$$\frac{1}{n}\mathrm{E}\ \boldsymbol{\nabla\nabla}' l_n\left(\alpha, \beta\right)\Big|_{(\alpha_0, \beta_0)} = -\mathbf{S}(\alpha_0, \beta_0) \text{ for } n = 1, \ldots\ .$$

$\mathbf{S}(\alpha_0, \beta_0)$ is shown to be nonsingular by evaluating the determinant of $\mathbf{S}(\alpha_0, \beta_0)$ when $|h(x)|$ is non-constant with respect to $g_2(x)$.

$$
\begin{aligned}
\text{let } \boldsymbol{M} &\equiv \frac{1+\rho_1}{\rho_1}\mathbf{S}(\alpha_0, \beta_0) \\
\det \boldsymbol{M} &= \mathrm{E}\left(\frac{h^2\left(X_1\right)}{D_1\left(X_1|\alpha_0,\beta_0\right)}\right)\mathrm{E}\left(\frac{1}{D_1\left(X_1|\alpha_0,\beta_0\right)}\right) - \mathrm{E}^2\left(\frac{h\left(X_1\right)}{D_1\left(X_1|\alpha_0,\beta_0\right)}\right) \\
&= \left(\mathrm{E}\left(h^2\left(X_*\right)\right) - \mathrm{E}^2\left(h\left(X_*\right)\right)\right)\mathrm{E}^2\left(\frac{1}{D_1\left(X_1|\alpha_0,\beta_0\right)}\right) \\
X_* \sim g_*\left(x\right) &= \mathrm{E}^{-1}\left(\frac{1}{D_1\left(X_1|\alpha_0,\beta_0\right)}\right)\frac{w_1\left(x|\alpha_0,\beta_0\right)}{D_1\left(x|\alpha_0,\beta_0\right)}g_2\left(x\right)
\end{aligned}
$$

Hence $\det \boldsymbol{M} = 0$ when $h(X_*)$ is a degenerate (variance 0) random variable, and $\det \boldsymbol{M} \neq 0$ when $|h(X_*)|$ is non-constant almost everywhere or equivalently when $|h(X_2)|$ is non-constant almost everywhere since $g_*(x)$ and $g_2(x)$ have the same support, see [6] equation 4.7.4 and Lemma 4.7.1,

With regard to condition (C), Lemma 2.8 will show (17). Equations (40), (41), and (43) show that

$$
\begin{aligned}
\mathbf{Var}\left(n^{-\frac{1}{2}}\ \boldsymbol{\nabla} l_n\left(\alpha,\beta\right)|_{(\alpha_0,\beta_0)}\right) &= \frac{\left(1+\rho_1\right)^2}{\rho_1}\mathbf{V}_0,\ n=1,2,\ldots \\
&= \boldsymbol{B}\left(\alpha_0,\beta_0\right)\ .
\end{aligned}
$$

With regard to condition (D), starting with (32) for $(\alpha, \beta) \in \boldsymbol{\Theta}$, applying a result (26) from Lemma 2.3 with parametric index $\boldsymbol{\Theta}$, and applying the law of large numbers, shows

$$
\begin{aligned}
\frac{1}{n}\boldsymbol{\nabla} l_n\left(\alpha,\beta\right) &\overset{as}{\rightarrow} \frac{\rho_1}{1+\rho_1}\begin{pmatrix} 1-\sum_{j=1}^2 \mathrm{E}\left(\frac{\rho_j w_1(X_j|\alpha,\beta)}{D_1(X_j|\alpha,\beta)}\right) \\ \mathrm{E}\left(h\left(X_1\right)\right)-\sum_{j=1}^2 \mathrm{E}\left(h\left(X_j\right)\frac{\rho_j w_1(X_j|\alpha,\beta)}{D_1(X_j|\alpha,\beta)}\right)\end{pmatrix} \\
&= \mathrm{E}\frac{1}{n}\boldsymbol{\nabla} l_n\left(\alpha,\beta\right) \equiv \boldsymbol{\nabla} g\left(\alpha,\beta\right)\ . \qquad (36)
\end{aligned}
$$

The following anti-derivative of $\boldsymbol{\nabla} g(\alpha, \beta)$ with respect to (α, β) is suggested, assuming the usual regularity conditions so that integration and differentiation may be interchanged

$$
\begin{aligned}
g\left(\alpha,\beta\right) &= \frac{1}{1+\rho_1}\left(\rho_1\left(\alpha+\beta\mathrm{E}\left(h\left(X_1\right)\right)\right)-\sum_{j=1}^{2}\rho_j\mathrm{E}\left(\log\left(D_1\left(X_j|\alpha,\beta\right)\right)\right)\right) \\
&= \mathrm{E}\frac{1}{n}l_n\left(\alpha,\beta\right)\ . \qquad (37)
\end{aligned}
$$

It will be shown that $n^{-1}l_n(\alpha,\beta)$ converges to $g(\alpha,\beta)$ almost surely uniformly in $(\alpha,\beta)\in\boldsymbol{\Theta}$ an open neighborhood of (α_0,β_0).

Definition 2.6. Let $\mathcal{F}_1(\boldsymbol{\Theta})$ and $\mathcal{F}_2(\boldsymbol{\Theta})$ denote two classes of functions, that are indexed by a bounded subset $\boldsymbol{\Theta}\subset\mathbb{R}^2$ containing (α_0,β_0), and that are integrable with respect to the probability distributions G_1 and G_2 associated with the densities g_1 and g_2, as defined by:

$$\mathcal{F}_1(\boldsymbol{\Theta})\equiv\{f_1(x|\alpha,\beta)=\log(D_1(x|\alpha,\beta))-(\alpha+\beta h(x)):(\alpha,\beta)\in\boldsymbol{\Theta}\}$$
$$\mathcal{F}_2(\boldsymbol{\Theta})\equiv\{f_2(x|\alpha,\beta)=\log(D_1(x|\alpha,\beta)):(\alpha,\beta)\in\boldsymbol{\Theta}\}$$

where $f_1\in L_1(G_1)$ and $f_2\in L_1(G_2)$.

The functions $f_1(x|\alpha,\beta)\in\mathcal{F}_1(\boldsymbol{\Theta})$ and $f_2(x|\alpha,\beta)\in\mathcal{F}_2(\boldsymbol{\Theta})$ have partial derivatives of all orders with respect to (α,β). A Taylor series expansion, for $f_1(x|\alpha,\beta)$ and for $f_2(x|\alpha,\beta)$ around $(\alpha,\beta)\in\boldsymbol{\Theta}$, and the mean value theorem 6.7 [16], identifies the following Lipschitz bound $m(x)$

$$\begin{aligned}
f_1\left(x|\alpha^1,\beta^1\right)-f_1\left(x|\alpha^2,\beta^2\right)&=\boldsymbol{\nabla}'f_1\left(x|\alpha_{\lambda^1},\beta_{\lambda^1}\right)\begin{pmatrix}\alpha^1-\alpha^2\\ \beta^1-\beta^2\end{pmatrix}\\
f_2\left(x|\alpha^1,\beta^1\right)-f_2\left(x|\alpha^2,\beta^2\right)&=\boldsymbol{\nabla}'f_2\left(x|\alpha_{\lambda^2},\beta_{\lambda^2}\right)\begin{pmatrix}\alpha^1-\alpha^2\\ \beta^1-\beta^2\end{pmatrix}\\
\left(\alpha_{\lambda^i},\beta_{\lambda^i}\right)&=\lambda^i\left(\alpha^1,\beta^1\right)+\left(1-\lambda^i\right)\left(\alpha^2,\beta^2\right),\ \lambda^i\in(0,1),\ i=1,2\\
\forall\ (\alpha,\beta)\in\boldsymbol{\Theta}:\left\|\boldsymbol{\nabla}'f_1\left(x|\alpha,\beta\right)\right\|_\infty&=\left\|-\frac{1}{D_1\left(x|\alpha,\beta\right)}\begin{pmatrix}1, & h\left(x\right)\end{pmatrix}\right\|_\infty\\
&\le\left(1+\left|h\left(x\right)\right|\right)\equiv m\left(x\right)\ ,\\
\forall\ (\alpha,\beta)\in\boldsymbol{\Theta}:\left\|\boldsymbol{\nabla}'f_2\left(x|\alpha,\beta\right)\right\|_\infty&=\left\|\frac{\rho_1 w_1\left(x|\alpha,\beta\right)}{D_1\left(x|\alpha,\beta\right)}\begin{pmatrix}1, & h\left(x\right)\end{pmatrix}\right\|_\infty\\
&\le m\left(x\right)\ .
\end{aligned}$$

Given any bounded subset $\boldsymbol{\Theta}\subset\mathbb{R}^2$, it is easy to show that any $f_j(x|\alpha,\beta)\in\mathcal{F}_j(\boldsymbol{\Theta})$ is integrable with respect G_j with $(\alpha,\beta)\in\boldsymbol{\Theta}$ and with $j=1,2$

$$\left|f_j\left(x|\alpha,\beta\right)\right|\le\left|f_j\left(x|0,0\right)\right|+m\left(x\right)\left\|(\alpha,\beta)\right\|_\infty\ .$$

Hence $f_j(x)\in L_1(G_j)$ and $m(x)\in L_1(G_j)$ for $j=1,2$ under the assumptions of this lemma. Applying Lemma 2.2 to $f_j(x|\alpha,\beta)\in\mathcal{F}_j(\boldsymbol{\Theta})$ for $j=1,2$ shows

$$\sup_{(\alpha,\beta)\in\boldsymbol{\Theta}}\left|\frac{1}{n_j}\sum_{i=1}^{n_j}f_j\left(x_{ji}|\alpha,\beta\right)-\mathrm{E}\left(f_j\left(X_j|\alpha,\beta\right)\right)\right|\overset{as*}{\to}0\ .$$

Given the following identities for $l_n(\alpha, \beta)$ and $g(\alpha, \beta)$

$$\begin{aligned}
\frac{1}{n} l_n(\alpha, \beta) &= \left(\frac{\rho_1}{1+\rho_1}\right)\left[\frac{1}{n_1}\sum_{i=1}^{n_1} \alpha + \beta h(x_{1i}) - \log\left(D_1(x_{1i}|\alpha, \beta)\right)\right] \\
&\quad - \left(\frac{1}{1+\rho_1}\right)\frac{1}{n_2}\sum_{i=1}^{n_2} \log\left(D_1(x_{2i}|\alpha, \beta)\right) \\
g(\alpha, \beta) &= \left(\frac{\rho_1}{1+\rho_1}\right)\left[\alpha + \beta \mathrm{E}\left(h(X_1)\right) - \mathrm{E}\left(\log\left(D_1(X_1|\alpha, \beta)\right)\right)\right] \\
&\quad - \left(\frac{1}{1+\rho_1}\right)\mathrm{E}\left(\log\left(D_1(X_2|\alpha, \beta)\right)\right)
\end{aligned}$$

then the combined result from the previous display shows that $n^{-1}l_n(\alpha, \beta)$ converges to $g(\alpha, \beta)$ almost surely uniformly in $(\alpha, \beta) \in \boldsymbol{\Theta}$.

$$\begin{aligned}
&\sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \left|\frac{1}{n} l_n(\alpha, \beta) - g(\alpha, \beta)\right| \\
&\leq \sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \frac{\rho_1}{1+\rho_1}\left|\frac{1}{n_1}\sum_{i=1}^{n_1} f_1(x_{1i}|\alpha, \beta) - \mathrm{E}\left(f_1(X_1|\alpha, \beta)\right)\right| \\
&+ \sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \frac{1}{1+\rho_1}\left|\frac{1}{n_2}\sum_{i=1}^{n_2} f_2(x_{2i}|\alpha, \beta) - \mathrm{E}\left(f_2(X_2|\alpha, \beta)\right)\right| \\
&\overset{as*}{\to} 0 \,.
\end{aligned}$$

Condition (D) is proven by specializing $\boldsymbol{\Theta}$ to an open bounded subset of $\mathbb{R}^2$ containing (α_0, β_0).

Condition (E) is proven by showing that the matrix $\boldsymbol{M}$ defined above is positive definite. Let $\boldsymbol{X} = (x_1, x_2)' \neq \boldsymbol{0}$.

$$\begin{aligned}
\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{bmatrix} A_0 & A_1 \\ A_1 & A_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= A_0 x_1^2 + 2A_1 x_1 x_2 + A_2 x_2^2 \\
&= \left(\sqrt{A_0} x_1 + \frac{A_1}{\sqrt{A_0}} x_2\right)^2 + \left(A_2 - \frac{A_1^2}{A_0}\right) x_2^2 \\
&= \left(\sqrt{A_0} x_1 + \frac{A_1}{\sqrt{A_0}} x_2\right)^2 + \frac{1}{A_0}\det(\boldsymbol{M})\, x_2^2
\end{aligned}$$

Hence $\boldsymbol{M}$ is positive definite if and only if $\det(\boldsymbol{M}) > 0$. So the result is proven when $|h(x)|$ is non-constant with respect to $g_2(x)$ resulting in $\det(\boldsymbol{M}) > 0$ as shown for condition (B) above.

For condition (F), the law of large numbers is applied to find the limit of $n^{-1}\boldsymbol{\nabla\nabla}' l_n(\alpha,\beta)$ for $(\alpha,\beta)\in\boldsymbol{\Theta}$. As a stronger result, the abstract Glivenko-Cantelli theorem is applied to find the limit of $n^{-1}\boldsymbol{\nabla\nabla}' l_n(\alpha,\beta)$ uniformly in $(\alpha,\beta)\in\boldsymbol{\Theta}$. For either application

$$\begin{aligned}
&\sum_{i=1}^{n} h^k\left(t_i\right)\hat{p}^2\left(t_i|\alpha,\beta\right) w_1\left(t_i|\alpha,\beta\right) n_1 \\
&\qquad\qquad \stackrel{as}{\to} \rho_1 \sum_{j=1}^{2} \mathrm{E}\left(h^k\left(X_j\right)\frac{\rho_j w_1\left(X_j|\alpha,\beta\right)}{D_1^2\left(X_j|\alpha,\beta\right)}\right) \\
&\qquad\qquad = \rho_1 \mathrm{E}\left(h^k\left(X_2\right)\frac{w_1\left(X_2|\alpha,\beta\right)}{D_1\left(X_2|\alpha,\beta\right)}\left(\frac{1+\rho_1 w_1\left(X_2|\alpha_0,\beta_0\right)}{D_1\left(X_2|\alpha,\beta\right)}\right)\right) \\
&\qquad\qquad \equiv \rho_1 A_k\left(\alpha,\beta\right),\ k=0,1,2 \\
\frac{1}{n}\boldsymbol{\nabla\nabla}' l_n\left(\alpha,\beta\right) &\stackrel{as}{\to} -\frac{\rho_1}{1+\rho_1}\begin{bmatrix} A_0\left(\alpha,\beta\right) & A_1\left(\alpha,\beta\right) \\ A_1\left(\alpha,\beta\right) & A_2\left(\alpha,\beta\right)\end{bmatrix} \\
&= \mathrm{E}\frac{1}{n}\boldsymbol{\nabla\nabla}' l_n\left(\alpha,\beta\right)\ . \qquad (38)
\end{aligned}$$

In summary, the six conditions (A) through (F) have been proven. Hence, one of the roots $(\alpha_*,\beta_*)\in\boldsymbol{\Theta}_n$ converges in probability to (α_0,β_0). ■

The previous extremum estimator analysis shows that one of the roots $(\alpha_*,\beta_*)\in\boldsymbol{\Theta}_n$ converges in probability to (α_0,β_0). If there are multiple local maximums of $g(\alpha,\beta)$ that satisfy the six conditions (A) through (F), then this analysis does not determine which one of the local maximums of $g(\alpha,\beta)$ is the limit in probability of $(\hat{\alpha},\hat{\beta})\in\boldsymbol{\Theta}_n$. To complete this analysis, it is shown that $g(\alpha,\beta)$ has a unique global maximum at (α_0,β_0) and that $(\hat{\alpha},\hat{\beta})$ converges in probability to (α_0,β_0).

Lemma 2.5. *Under the conditions of Lemma 2.4, if $h(x)$ is continuous then $g(\alpha,\beta)$ has a unique global maximum at (α_0,β_0).*

Proof: Let $\boldsymbol{\Theta}$ denote a bounded subset of $\mathbb{R}^2$ that contains two local maximums (α_0,β_0) and (α_1,β_1) of $g(\alpha,\beta)$, i.e. $(\alpha_0,\beta_0),(\alpha_1,\beta_1)\in\boldsymbol{\Theta}$. Starting with (32) with $(\alpha,\beta)=(\alpha_*,\beta_*)\in\boldsymbol{\Theta}_n$ and applying the convergence property (23) of Lemma 2.3 to the classes of functions $\mathcal{F}^{(1)}_{j|k}(\boldsymbol{\Theta})\equiv\mathcal{F}_j(h^k(x),1)$ with parametric index $\boldsymbol{\Theta}$ for $j=1,2$ and $k=0,1$ where $(\alpha_*,\beta_*)\in\boldsymbol{\Theta}_n\stackrel{P}{\to}(\alpha_1,\beta_1)\in\boldsymbol{\Theta}$, and where $(\hat{\alpha},\hat{\beta})\in\boldsymbol{\Theta}_n\stackrel{P}{\to}(\alpha_0,\beta_0)\in\boldsymbol{\Theta}$, shows that (α_*,β_*)

and $(\hat{\alpha}, \hat{\beta})$ are zeros of the function $\boldsymbol{\nabla} g(\alpha, \beta)$

$$\begin{aligned}
\mathbf{0} &= \frac{1}{n} \boldsymbol{\nabla} l_n \left(\alpha_*, \beta_*\right) \xrightarrow{P} \boldsymbol{\nabla} g \left(\alpha_1, \beta_1\right) \\
\mathbf{0} &= \frac{1}{n} \boldsymbol{\nabla} l_n \left(\hat{\alpha}, \hat{\beta}\right) \xrightarrow{P} \boldsymbol{\nabla} g \left(\alpha_0, \beta_0\right) \ .
\end{aligned}$$

After a little algebra, the previous display is rewritten as

$$\begin{aligned}
\kappa \equiv \mathrm{E}\left(\frac{w_1\left(X_2|\alpha_0,\beta_0\right)}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right) &= \mathrm{E}\left(\frac{w_1\left(X_2|\alpha_1,\beta_1\right)}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right) \\
\mu_h^* \equiv \mathrm{E}\left(\frac{h\left(X_2\right)}{\kappa}\frac{w_1\left(X_2|\alpha_0,\beta_0\right)}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right) &= \mathrm{E}\left(\frac{h\left(X_2\right)}{\kappa}\frac{w_1\left(X_2|\alpha_1,\beta_1\right)}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right) \\
0 = \mathrm{E}\left(\frac{\left(h\left(X_2\right)-\mu_h^*\right)}{\kappa}\frac{w_1\left(X_2|\alpha_0,\beta_0\right)}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right)& \\
= \mathrm{E}\left(\frac{\left(h\left(X_2\right)-\mu_h^*\right)}{\kappa}\frac{w_1\left(X_2|\alpha_1,\beta_1\right)}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right)& \\
0 = \mathrm{E}\left(\left(h\left(X_2\right)-\mu_h^*\right)\frac{e^{\beta_0\left(h(X_2)-\mu_h^*\right)}}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right)& \\
= \mathrm{E}\left(\left(h\left(X_2\right)-\mu_h^*\right)\frac{e^{\beta_1\left(h(X_2)-\mu_h^*\right)}}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right)& \ .
\end{aligned}$$

It is easy to show for $x \in \{x : h(x) - \mu_h^* \neq 0\}$ and $\beta_0 < \beta_1$ that

$$\left(h\left(x\right)-\mu_h^*\right) e^{\beta_0\left(h(x)-\mu_h^*\right)} < \left(h\left(x\right)-\mu_h^*\right) e^{\beta_1\left(h(x)-\mu_h^*\right)} \ .$$

Using the previous display and assuming $h(x)$ is continuous and non-constant with respect to $g(x)$ results in

$$\mathrm{E}\left(\left(h\left(X_2\right)-\mu_h^*\right)\frac{e^{\beta_0\left(h(X_2)-\mu_h^*\right)}}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right) < \mathrm{E}\left(\left(h\left(X_2\right)-\mu_h^*\right)\frac{e^{\beta_1\left(h(X_2)-\mu_h^*\right)}}{D_1\left(X_2|\alpha_1,\beta_1\right)}\right)$$

implying that $\beta_1 \leq \beta_0$. A similar analysis for $\beta_1 < \beta_0$ implies that $\beta_0 \leq \beta_1$. Hence there exist a single zero (α_0, β_0) of the function $\boldsymbol{\nabla} g(\alpha, \beta)$ implying a unique global maximum (α_0, β_0) of the function $g(\alpha, \beta)$.

As an alternate proof of $g(\alpha, \beta)$ having a global maximum at (α_0, β_0), $\boldsymbol{\nabla} g(\alpha, \beta)$ is shown to equal zero at (α_0, β_0) and $\boldsymbol{\nabla}\boldsymbol{\nabla}' g(\alpha, \beta)$ is shown to be negative definite for all $(\alpha, \beta) \in \mathbb{R}^2$. Using the following bounds on the first and second partial derivatives of $n^{-1}l_n(\alpha,\beta)$,

$$\left|\frac{\partial}{\partial\alpha}\frac{l(\alpha,\beta)}{n}\right| \leq 1,\ \left|\frac{\partial}{\partial\beta}\frac{l(\alpha,\beta)}{n}\right| \leq \frac{1}{n}\sum_{t=1}^{n}\left|h\left(x_t\right)\right|$$

$$\left|\frac{\partial^2}{\partial\alpha^2}\frac{l(\alpha,\beta)}{n}\right| \leq 1,\ \left|\frac{\partial^2}{\partial\alpha\partial\beta}\frac{l(\alpha,\beta)}{n}\right| \leq \frac{1}{n}\sum_{t=1}^{n}\left|h\left(x_t\right)\right|,$$

$$\left|\frac{\partial^2}{\partial\beta^2}\frac{l(\alpha,\beta)}{n}\right| \leq \frac{1}{n}\sum_{t=1}^{n}h^2\left(x_t\right)$$

and using Corollary 2.4.1 of Theorem 2.4.2 from [6], shows

$$\boldsymbol{\nabla}g\left(\alpha,\beta\right) = \mathrm{E}\frac{1}{n}\boldsymbol{\nabla}l_n\left(\alpha,\beta\right),\ \boldsymbol{\nabla}\boldsymbol{\nabla}'g\left(\alpha,\beta\right) = \mathrm{E}\frac{1}{n}\boldsymbol{\nabla}\boldsymbol{\nabla}'l_n\left(\alpha,\beta\right)$$

where $h^k(x)$ for $k=1,2$ is assumed to be integrable with respect to $g_j(x)$ for $j=1,2$. The structure of $\boldsymbol{\nabla}g(\alpha,\beta)$ from (36) implies that $\boldsymbol{\nabla}g(\alpha_0,\beta_0)=\mathbf{0}$. The structure of $\boldsymbol{\nabla}\boldsymbol{\nabla}'g(\alpha,\beta)$ from (38) implies that $-\boldsymbol{\nabla}\boldsymbol{\nabla}'g(\alpha,\beta)$ is positive definite for all $(\alpha,\beta)\in\mathbb{R}^2$ if and only if the determinant of $-\boldsymbol{\nabla}\boldsymbol{\nabla}'g(\alpha,\beta)$ is positive for all $(\alpha,\beta)\in\mathbb{R}^2$. The Cauchy-Schwarz inequality shows

$$\det\left(-\frac{1+\rho_1}{\rho_1}\boldsymbol{\nabla}\boldsymbol{\nabla}'g(\alpha,\beta)\right) = A_2\left(\alpha,\beta\right)A_0\left(\alpha,\beta\right) - A_1^2\left(\alpha,\beta\right) \geq 0\ .$$

The determinant equals 0 if and only if $|h(x)|$ is constant almost everywhere with respect to $g_2(x)$. Hence under the assumptions of this lemma, $\boldsymbol{\nabla}\boldsymbol{\nabla}'g(\alpha,\beta)$ is negative definite for all $(\alpha,\beta)\in\mathbb{R}^2$. Thus with $\boldsymbol{\nabla}g(\alpha_0,\beta_0)=\mathbf{0}$, a second order Taylor series expansion of $g(\alpha,\beta)$ around (α_0,β_0) shows that $g(\alpha,\beta)$ has a global maximum at (α_0,β_0). ■

Lemma 2.6. *Let $\boldsymbol{\Theta}$ denote a bounded subset of $\mathbb{R}^2$ that contains (α_0,β_0) as an interior point. If $n^{-1}l_n(\alpha,\beta)$ converges uniformly in probability to $g(\alpha,\beta)$ for $(\alpha,\beta)\in\boldsymbol{\Theta}$ where $g(\alpha,\beta)$ has a global maximum at (α_0,β_0) and if $h(x)$ is non-constant with respect to $g_2(x)$ then $(\hat{\alpha},\hat{\beta})$ converges in probability to (α_0,β_0).*

Proof: Let $\boldsymbol{\Theta}_0$ denote a closed bounded subset of $\boldsymbol{\Theta}$ that contains (α_0,β_0) as an interior point and that contains the boundary of $\boldsymbol{\Theta}_0$ denoted as $\partial(\boldsymbol{\Theta}_0)$. Using the assumption that $n^{-1}l_n(\alpha,\beta)$ converges uniformly in probability to $g(\alpha,\beta)$ for $(\alpha,\beta)\in\boldsymbol{\Theta}$, and using the assumption that $g(\alpha,\beta)$ has a unique global maximum at (α_0,β_0), shows that

$$P\left(\frac{1}{n}l_n\left(\alpha_0,\beta_0\right) > \sup_{(\alpha,\beta)\in\partial(\boldsymbol{\Theta}_0)}\frac{1}{n}l_n\left(\alpha,\beta\right)\right) \to 1\ .$$

The set in the previous display implies the existence of a local maximum for $n^{-1}l_n(\alpha,\beta)$ at (α^*,β^*) in the interior of $\mathbf{\Theta}_0$.

The determinant of $n^{-1}\boldsymbol{\nabla}\boldsymbol{\nabla}' l_n(\alpha,\beta)$ is shown to be greater than or equal to 0 by applying the Cauchy-Schwarz inequality for vectors (identified as inequality 1e.1) from [23]. The singular condition occurs if and only if $h(t_i)$ is constant for all $i = 1,\ldots,n$. Hence $n^{-1}\boldsymbol{\nabla}\boldsymbol{\nabla}' l_n(\alpha,\beta)$ is negative definite almost surely under the assumptions of this lemma. A second order Taylor series expansion of $n^{-1}\boldsymbol{\nabla}\boldsymbol{\nabla}' l_n(\alpha,\beta)$ about (α^*,β^*) shows that there exists a single global maximum of $n^{-1}l_n(\alpha,\beta)$ at $(\hat{\alpha},\hat{\beta})$ almost surely.

Hence the result is proven since the existence of a single local maximum almost surely such that $(\alpha^*,\beta^*) = (\hat{\alpha},\hat{\beta})$ shows

$$P\left(\frac{1}{n}l_n(\alpha_0,\beta_0) > \sup_{(\alpha,\beta)\in\partial(\mathbf{\Theta}_0)} \frac{1}{n}l_n(\alpha,\beta)\right) \leq P\left(\left(\hat{\alpha},\hat{\beta}\right)\in\mathbf{\Theta}_0\right) \to 1. \blacksquare$$

Corollary 2.9. *Under the conditions of Lemma 2.6, if* $(\hat{\alpha},\hat{\beta}) \xrightarrow{P} (\alpha_0,\beta_0)$ *and* $(\alpha_*,\beta_*) = \lambda(\hat{\alpha},\hat{\beta}) + (1-\lambda)(\alpha_0,\beta_0)$ *for some* $\lambda \in (0,1)$ *then* $(\alpha_*,\beta_*) \xrightarrow{P} (\alpha_0,\beta_0)$*.*

Proof: The result is proven by letting $\mathbf{\Theta}$ denote any open bounded convex subset of $\mathbb{R}^2$ containing (α_0,β_0) and applying Lemma 2.6 to show

$$P\left(\left(\hat{\alpha},\hat{\beta}\right)\in\mathbf{\Theta}\right) \leq P\left((\alpha_*,\beta_*)\in\mathbf{\Theta}\right) \to 1. \blacksquare$$

Corollary 2.10. *Under the conditions of Lemma 2.6, if* $(\hat{\alpha},\hat{\beta}) \xrightarrow{P} (\alpha_0,\beta_0)$ *then applying Corollary 2.9 to (8) and (9) shows the convergence in probability of* $(\acute{\alpha},\acute{\beta})$ *and* $(\grave{\alpha},\grave{\beta})$ *to* (α_0,β_0)*.* $\blacksquare$

The following display summarizes the convergence results proved above

$$\begin{aligned}
\left(\hat{\alpha},\hat{\beta}\right) &\xrightarrow{P} (\alpha_0,\beta_0) && \text{from (12)}\\
\hat{\mu}_h(\alpha_0,\beta_0) &\xrightarrow{as} \mu_h && \text{from (13)}\\
\hat{\sigma}_h^2\left(\hat{\alpha},\hat{\beta}\right) &\xrightarrow{P} \sigma_h^2 && \text{from (14)}\\
\boldsymbol{\nabla}\hat{\sigma}_h^2(\alpha,\beta)\big|_{(\acute{\alpha},\acute{\beta})} &\xrightarrow{P} \boldsymbol{\nabla}\sigma_h^2(\alpha_0,\beta_0) && \text{from (15)}\\
-\frac{1}{n}\,\boldsymbol{\nabla}\boldsymbol{\nabla}' l(\alpha,\beta)\big|_{(\grave{\alpha},\grave{\beta})} &\xrightarrow{P} \mathbf{S}(\alpha_0,\beta_0) && \text{from (16)}
\end{aligned}$$

Lemma 2.7. *Under the convergence conditions defined in Assumption 2.1, $\boldsymbol{D}_n$ from (10) converges in probability to $\boldsymbol{D} = \boldsymbol{D}(\alpha_0, \beta_0)$ as follows*

$$\boldsymbol{D}_n \xrightarrow{P} \boldsymbol{D}(\alpha_0,\beta_0) = \frac{1}{2\sigma_h}\left(-2\mu_h\beta_0, \quad \beta_0, \quad \mathbf{Q}(\alpha_0,\beta_0)\,\mathbf{S}^{-1}(\alpha_0,\beta_0)\right)' \,. \tag{39}$$

Proof: The continuous mapping theorem, Slutsky's theorem, and Corollaries 2.1, 2.7, and 2.8 are applied to prove the result that $\boldsymbol{D}_n \xrightarrow{P} \boldsymbol{D}$. ■

Remark 2.3. Next the asymptotic distribution is shown for $\boldsymbol{Y}_n$, as previously defined in (11), using the following decomposition

$$\boldsymbol{Y}_n = \begin{pmatrix} Y_{1n} \\ Y_{2n} \\ Y_{3n} \\ Y_{4n} \end{pmatrix} \equiv \sqrt{\frac{n_1 n_2}{n}} \begin{pmatrix} \hat{\mu}_h(\alpha_0,\beta_0) - \mu_h \\ \hat{\mu}_{h^2}(\alpha_0,\beta_0) - \mu_{h^2} \\ \frac{1}{n}\,\boldsymbol{\nabla} l(\alpha,\beta)\big|_{(\alpha_0,\beta_0)} \end{pmatrix} \tag{40}$$

$$\equiv \frac{1}{\sqrt{n_1}}\sum_{i=1}^{n_1}\left(\boldsymbol{Y}_{1i} - \mathbf{E}(\boldsymbol{Y}_1)\right) + \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2}\left(\boldsymbol{Y}_{2i} - \mathbf{E}(\boldsymbol{Y}_2)\right)$$

$$\boldsymbol{Y}_{1i} = \boldsymbol{M}_1 \begin{pmatrix} \frac{h(x_{1i})}{D_1(x_{1i}|\alpha_0,\beta_0)} \\ \frac{h^2(x_{1i})}{D_1(x_{1i}|\alpha_0,\beta_0)} \\ \frac{1}{D_1(x_{1i}|\alpha_0,\beta_0)} \\ \frac{h(x_{1i})}{D_1(x_{1i}|\alpha_0,\beta_0)} \end{pmatrix}, \; \boldsymbol{M}_1 = \sqrt{\frac{1}{1+\rho_1}} \begin{bmatrix} \rho_1 & & & \\ & \rho_1 & & \\ & & \frac{\rho_1}{1+\rho_1} & \\ & & & \frac{\rho_1}{1+\rho_1} \end{bmatrix}$$

$$\boldsymbol{Y}_{2i} = \boldsymbol{M}_2 \begin{pmatrix} \frac{h(x_{2i})}{D_1(x_{2i}|\alpha_0,\beta_0)} \\ \frac{h^2(x_{2i})}{D_1(x_{2i}|\alpha_0,\beta_0)} \\ \frac{1}{D_1(x_{2i}|\alpha_0,\beta_0)} \\ \frac{h(x_{2i})w_1(x_{2i}|\alpha_0,\beta_0)}{D_1(x_{2i}|\alpha_0,\beta_0)} \end{pmatrix}, \; \boldsymbol{M}_2 = \sqrt{\frac{\rho_1}{1+\rho_1}} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \frac{1}{1+\rho_1} & \\ & & & -\frac{\rho_1}{1+\rho_1} \end{bmatrix}$$

$$\boldsymbol{Y}_{1i} \sim \left(\mathbf{E}(\boldsymbol{Y}_1), \mathbf{Var}(\boldsymbol{Y}_1)\right), \; i = 1,\ldots,n_1$$
$$\boldsymbol{Y}_{2i} \sim \left(\mathbf{E}(\boldsymbol{Y}_2), \mathbf{Var}(\boldsymbol{Y}_2)\right), \; i = 1,\ldots,n_2 \,.$$

Notice that $\mathrm{E}(\hat{\mu}_{h^k}(\alpha_0,\beta_0)) = \mu_{h^k} \equiv \mathrm{E}(h(X_2))$ for $k = 1,2$ where $\hat{\mu}_{h^k}(\alpha_0,\beta_0)$ depends on (α_0,β_0) but μ_{h^k} does not depend on (α_0,β_0). This is true because $\hat{\mu}_{h^k}(\alpha_0,\beta_0)$ consists of two random samples from X_1 and X_2 with means that satisfy

$$\hat{\mu}_{h^k}(\alpha_0,\beta_0) = \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\rho_1 h^k(x_{1i})}{D_1(x_{1i}|\alpha_0,\beta_0)} + \frac{1}{n_2}\sum_{i=1}^{n_2}\frac{h^k(x_{2i})}{D_1(x_{2i}|\alpha_0,\beta_0)}$$

$$\mathrm{E}\left(\hat{\mu}_{h^k}(\alpha_0,\beta_0)\right) = \mathrm{E}\left(\frac{\rho_1 h^k(X_1)}{D_1(X_1|\alpha_0,\beta_0)}\right) + \mathrm{E}\left(\frac{h^k(X_2)}{D_1(X_2|\alpha_0,\beta_0)}\right) = \mathrm{E}\left(h^k(X_2)\right)$$

where the individual means depend on (α_0, β_0) but the sum of the means does not depend on (α_0, β_0).

Lemma 2.8. *Assuming $h^k(x)$ is square integrable for $k = 0, 1, 2$ with respect to $g_1(x)$ and $g_2(x)$, then $\boldsymbol{Y}_n$ converges in distribution to a multivariate Gaussian distribution $\boldsymbol{Y}$:*

$$\begin{gathered}\boldsymbol{Y}_n = (Y_{1n}, Y_{2n}, Y_{3n}, Y_{4n})' \xrightarrow{d} \boldsymbol{Y} = (Y_1, Y_2, Y_3, Y_4)' \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}) \\ \boldsymbol{\Sigma}_n \equiv \mathbf{Var}\left(\boldsymbol{Y}_n\right) = \mathbf{Var}\left(\boldsymbol{Y}_1\right) + \mathbf{Var}\left(\boldsymbol{Y}_2\right) = \boldsymbol{\Sigma}\ .\end{gathered} \tag{41}$$

Proof: The multivariate central limit theorem ([23], 2c.5) is applied to show the convergence in joint distribution of $\boldsymbol{Y}_n$ by showing every linear combination of $\boldsymbol{Y}_n$ converges in distribution to a univariate Gaussian distribution

$$\begin{gathered}z_n = \boldsymbol{\lambda}'\boldsymbol{Y}_n \xrightarrow{d} z = \boldsymbol{\lambda}'\boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}\right) \\ \boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)'\ .\end{gathered} \tag{42}$$

The Lindeberg-Feller form of the central limit theorem ([23], 2c.5) is applied to show (42).

$$\begin{aligned}
\text{Let } z_{ji} &\equiv \frac{1}{\sqrt{\rho_j}}\boldsymbol{\lambda}'\left(\boldsymbol{Y}_{ji} - \mathrm{E}\left(\boldsymbol{Y}_j\right)\right) \sim G_{z_{ji}} = G_{Z_j},\ j = 1, 2,\ i = 1, \ldots, n_j \\
Z_j &\sim \left(\mathrm{E}\left(Z_j\right), \mathrm{Var}\left(Z_j\right)\right) = \left(0, \frac{1}{\rho_j}\boldsymbol{\lambda}'\mathbf{Var}\left(\boldsymbol{Y}_j\right)\boldsymbol{\lambda}\right),\ j = 1, 2 \\
\text{Let } C_n^2 &\equiv \sum_{i=1}^{n_1}\mathrm{Var}\left(z_{1i}\right) + \sum_{i=1}^{n_2}\mathrm{Var}\left(z_{2i}\right) \\
&= \frac{n_1}{\rho_1}\boldsymbol{\lambda}'\mathbf{Var}\left(\boldsymbol{Y}_1\right)\boldsymbol{\lambda} + n_2\boldsymbol{\lambda}'\mathbf{Var}\left(\boldsymbol{Y}_2\right)\boldsymbol{\lambda} \\
&= n_2\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}
\end{aligned}$$

The Lindeberg-Feller convergence condition, as specialized to (42), is satisfied for any $\varepsilon > 0$

$$\begin{aligned}
&\frac{1}{C_n^2}\left(\sum_{i=1}^{n_1}\int_{|z|>\varepsilon C_n} z^2 dG_{z_{1i}}(z) + \sum_{i=1}^{n_2}\int_{|z|>\varepsilon C_n} z^2 dG_{z_{2i}}(z)\right) \\
=&\frac{\rho_1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}}\int I\left(|z| > \varepsilon C_n\right) z^2 dG_{Z_1}(z) + \frac{1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}}\int I\left(|z| > \varepsilon C_n\right) z^2 dG_{Z_2}(z) \\
&\rightarrow 0 \text{ as } n \uparrow \infty
\end{aligned}$$

since $\mathrm{Var}(z_n) = \boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda} = \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}$ is constant and finite for all n and since the convergence of the two integrals to zero follows by applying the dominated convergence theorem, hence

$$\frac{\sum_{i=1}^{n_1} z_{1i} + \sum_{i=1}^{n_2} z_{2i}}{\sqrt{n_2 \boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}}} \xrightarrow{d} \mathrm{N}\left(0, 1\right)$$

which proves the result that

$$\begin{aligned}\boldsymbol{\lambda}'\boldsymbol{Y}_n &= \frac{\sqrt{\rho_1}}{\sqrt{n_1}} \sum_{i=1}^{n_1} z_{1i} + \frac{1}{\sqrt{n_2}} \sum_{i=1}^{n_2} z_{2i} = \frac{1}{\sqrt{n_2}} \left(\sum_{i=1}^{n_1} z_{1i} + \sum_{i=1}^{n_2} z_{2i} \right) \\ &\xrightarrow{d} \mathrm{N}\left(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}\right) . \blacksquare\end{aligned}$$

In order to calculate $\mathbf{Var}(\boldsymbol{Y}_1)$ and $\mathbf{Var}(\boldsymbol{Y}_2)$, the following definitions are useful for $k = 0, \ldots, 4$, for $i = 0, 1, 2$, and for $j = 0, 1, 2$

$$\begin{aligned}
A_k &\equiv \mathrm{E}\left(\frac{h^k(X_1)}{D_1(X_1|\alpha_0, \beta_0)} \right), \; B_k \equiv \mathrm{E}\left(\frac{h^k(X_2)}{D_1(X_2|\alpha_0, \beta_0)} \right), \\
A_{ij} &\equiv \mathrm{E}\left(\frac{h^i(X_1)}{D_1(X_1|\alpha_0, \beta_0)} - A_i \right) \left(\frac{h^j(X_1)}{D_1(X_1|\alpha_0, \beta_0)} - A_j \right), \\
B_{ij} &\equiv \mathrm{E}\left(\frac{h^i(X_2)}{D_1(X_2|\alpha_0, \beta_0)} - B_i \right) \left(\frac{h^j(X_2)}{D_1(X_2|\alpha_0, \beta_0)} - B_j \right), \\
C_{ij} &\equiv \mathrm{E}\left(\frac{h^i(X_2)}{D_1(X_2|\alpha_0, \beta_0)} - B_i \right) \left(h^j(X_2) \frac{w_1(X_2|\alpha_0, \beta_0)}{D_1(X_2|\alpha_0, \beta_0)} - A_j \right), \\
D_2 &\equiv \mathrm{E}\left(h(X_2) \frac{w_1(X_2|\alpha_0, \beta_0)}{D_1(X_2|\alpha_0, \beta_0)} - B_1 \right)^2 .
\end{aligned}$$

The resulting expressions for $\mathbf{Var}(\boldsymbol{Y}_1)$ and $\mathbf{Var}(\boldsymbol{Y}_2)$ are

$$\mathbf{Var}\left(\boldsymbol{Y}_1\right) = \boldsymbol{M}_1 \begin{bmatrix} A_{11} & A_{12} & A_{10} & A_{11} \\ A_{21} & A_{22} & A_{20} & A_{21} \\ A_{01} & A_{02} & A_{00} & A_{01} \\ A_{11} & A_{12} & A_{10} & A_{11} \end{bmatrix} \boldsymbol{M}_1$$

$$\mathbf{Var}\left(\boldsymbol{Y}_2\right) = \boldsymbol{M}_2 \begin{bmatrix} B_{11} & B_{12} & B_{10} & C_{11} \\ B_{21} & B_{22} & B_{20} & C_{21} \\ B_{01} & B_{02} & B_{00} & C_{01} \\ C_{11} & C_{21} & C_{01} & D_2 \end{bmatrix} \boldsymbol{M}_2 .$$

A little algebra is used to simplify $\mathbf{\Sigma}_n = \mathbf{Var}(\boldsymbol{Y}_1) + \mathbf{Var}(\boldsymbol{Y}_2)$

$$\text{Let } \mathbf{\Sigma}_n \equiv \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_2 \\ \mathbf{\Sigma}_2' & \mathbf{V}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{Var}\begin{pmatrix} Y_{1n} \\ Y_{2n} \end{pmatrix} & \mathbf{Cov}\begin{pmatrix} Y_{1n} \\ Y_{2n} \end{pmatrix}\begin{pmatrix} Y_{3n} \\ Y_{4n} \end{pmatrix} \\ \mathbf{Cov}\begin{pmatrix} Y_{3n} \\ Y_{4n} \end{pmatrix}\begin{pmatrix} Y_{1n} \\ Y_{2n} \end{pmatrix} & \mathbf{Var}\begin{pmatrix} Y_{3n} \\ Y_{4n} \end{pmatrix} \end{bmatrix}$$

$$\begin{aligned}
\mathbf{\Sigma}_1 &= \frac{\rho_1}{1+\rho_1} \begin{bmatrix} \left(B_2 - B_1^2 - \rho_1 A_1^2\right) & \left(B_3 - B_1 B_2 - \rho_1 A_1 A_2\right) \\ \left(B_3 - B_1 B_2 - \rho_1 A_1 A_2\right) & \left(B_4 - B_2^2 - \rho_1 A_2^2\right) \end{bmatrix} \\
\mathbf{V}_0 &= \frac{\rho_1^2}{(1+\rho_1)^2} \begin{bmatrix} \left(\frac{1}{1+\rho_1} A_0 - A_0^2\right) & \left(\frac{1}{1+\rho_1} A_1 - A_0 A_1\right) \\ \left(\frac{1}{1+\rho_1} A_1 - A_0 A_1\right) & \left(\frac{1}{1+\rho_1} A_2 - A_1^2\right) \end{bmatrix} \qquad (43) \\
\mathbf{\Sigma}_2 &= \frac{\rho_1^2}{(1+\rho_1)^2} \begin{bmatrix} \left(B_1 - A_1\right) A_0 & \left(B_1 - A_1\right) A_1 \\ \left(B_2 - A_2\right) A_0 & \left(B_2 - A_2\right) A_1 \end{bmatrix} .
\end{aligned}$$

Theorem 2.2. *Under the convergence conditions identified in Assumption 2.1, $\tilde{Z}_n^*$ converges to a Gaussian random variable $\tilde{Z}^*$.*

Proof: The convergence in distribution of $\tilde{Z}_n^*$ as $n \to \infty$ is established using Slutsky's theorem, Lemma 2.7, and Lemma 2.8

$$\tilde{Z}_n^* = \boldsymbol{D}_n' \boldsymbol{Y}_n \xrightarrow{d} \tilde{Z}^* = \boldsymbol{D}' \boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{D}' \mathbf{\Sigma} \boldsymbol{D}\right) . \blacksquare \qquad (44)$$

The matrix algebra of $\boldsymbol{D}'\mathbf{\Sigma}\boldsymbol{D}$ is simplified by taking advantage of the structure of $\mathbf{S}(\alpha_0, \beta_0)$ in order to define

$$\begin{aligned}
\mathbf{S} \equiv \mathbf{S}\left(\alpha_0, \beta_0\right) &= \frac{\rho_1}{1+\rho_1} \begin{bmatrix} A_0 & A_1 \\ A_1 & A_2 \end{bmatrix} \\
\boldsymbol{M} &\equiv \begin{bmatrix} \mathbf{I}_2 & \\ & \mathbf{S}^{-1} \end{bmatrix} \text{ so that } \boldsymbol{M}\boldsymbol{Y} \sim \mathrm{E}\left(\mathbf{0}, \boldsymbol{M}\mathbf{\Sigma}\boldsymbol{M}\right) \\
\begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_3 \\ \mathbf{\Sigma}_3' & \mathbf{V}_1 \end{bmatrix} \equiv \boldsymbol{M}\mathbf{\Sigma}\boldsymbol{M} &= \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_2 \mathbf{S}^{-1} \\ \mathbf{S}^{-1}\mathbf{\Sigma}_2' & \mathbf{S}^{-1}\mathbf{V}_0\mathbf{S}^{-1} \end{bmatrix} \\
\mathbf{V}_1 &= \frac{1}{1+\rho_1} \begin{bmatrix} A_0 & A_1 \\ A_1 & A_2 \end{bmatrix}^{-1} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\
\mathbf{\Sigma}_3 &= \frac{\rho_1}{1+\rho_1} \begin{bmatrix} (B_1 - A_1) & 0 \\ (B_2 - A_2) & 0 \end{bmatrix}
\end{aligned}$$

and then rewritting the distribution of the random variable $\tilde{\mathrm{Z}}^*$ as

$$\begin{aligned}
\tilde{\mathrm{Z}}^* &= \boldsymbol{D}_1'\boldsymbol{MY} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{D}_1'\boldsymbol{M\Sigma M D}_1\right) \\
\boldsymbol{D}_1 &\equiv \frac{1}{2\sigma_h}\left(-2\mu_h\beta_0, \quad \beta_0, \quad \mathbf{Q}\left(\alpha_0, \beta_0\right)\right)' \\
\boldsymbol{MY} &= \left(Y_1, \quad Y_2, \quad Y_{\alpha_0}, \quad Y_{\beta_0}\right)' \\
\begin{pmatrix} Y_{\alpha_0} \\ Y_{\beta_0} \end{pmatrix} &\sim \frac{\sqrt{\rho_1}}{(1+\rho_1)}\begin{pmatrix} Z_{\alpha_0} \\ Z_{\beta_0} \end{pmatrix}; \text{ see (4) and (5).}
\end{aligned}$$

Under the alternative hypothesis, $\mathbf{H}_1 : \beta_0 \neq 0$ with β_0 fixed, Theorem 2.2 shows an asymptotic Gaussian distribution result

$$\tilde{\mathrm{Z}}_n^* = \sqrt{\frac{n_1 n_2}{n}}\left(\hat{\sigma}_h(\hat{\alpha}, \hat{\beta})\hat{\beta} - \sigma_h\beta_0\right) = \boldsymbol{D}_n'\boldsymbol{Y}_n \xrightarrow{d} \tilde{\mathrm{Z}}^* = \boldsymbol{D}'\boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{D}'\boldsymbol{\Sigma D}\right) .$$

This asymptotic Gaussian distribution will be used in section 2.2.3 in order to approximate the relative efficiency of the t-test to the semiparametric test. Section 2.2.3 also describes another type of efficiency called Pitman efficiency. To justify using this asymptotic Gaussian distribution in order to approximate the Pitman efficiency the following convergence in distribution result, a generalization of Theorem 2.2, is also needed

$$\tilde{\mathrm{Z}}_n^* = \sqrt{\frac{n_1 n_2}{n}}\left(\hat{\sigma}_h(\hat{\alpha}, \hat{\beta})\hat{\beta} - \sigma_h\beta_n\right) \xrightarrow{d(\beta_n)} \tilde{\mathrm{Z}}^* \sim \mathrm{N}\left(0, 1\right)$$

where the true distortion parameter β_n at time index n represents a sequence of alternative hypotheses, $\mathbf{H}_1 : \beta_n \neq 0$, such that $\beta_n \to \beta_0 = 0$. In general the results of Theorem 2.2 for any fixed $\beta_0 \neq 0$ do not imply the previous display.

Assumption 2.2. The following list defines convergence conditions that allow $\tilde{\mathrm{Z}}_n^*$ to converge to a Gaussian random variable $\tilde{\mathrm{Z}}^*$ as the true distortion parameter β_n converges to β_0:

- The random variable X_1 is distributed according to a sequence of density functions $\{p_n(x) : n = 1, 2, \dots\}$ where $X_1 \sim g_1 = p_n$ at time index n such that $p_n \to p_0$ almost everywhere where $p_0(x)$ defines another density function.
- The random variable X_2 is distributed according to the density function g_2 at all time indexes n: $X_2 \sim g_2$.

- The sequence of distortion parameters (α_n, β_n) converges to the limiting distortion parameters (α_0, β_0) where the density ratios $p_n(x)/g_2(x) = \exp(\alpha_n + \beta_n h(x))$ identify (α_n, β_n) and where the limiting density ratio $p_0(x)/g_2(x) = \exp(\alpha_0 + \beta_0 h(x))$ identifies (α_0, β_0).

- $h(x)$ is continuous and non-constant with respect to the density g_2 such that $P_{g_2}(x : h(x) = m) = 0$ for all $m \in \mathbb{R}$.

- $h^k(x)$ is integrable with respect to the sequence of densities $\{g_2, p_n : n = 0, 1, 2, \dots\}$ for $k = 1, 2, 3, 4$ such that $E_n|h^k(X_1)| \to E_0|h^k(X_1)|$ where the E_n notation denotes expectation according to the p_n density.

For the last convergence condition, $|h^k(x)|$ is bounded by $1 + h^4(x)$ for $k \in \{1, 2, 3\}$. If $E_n h^4(X_1) \to E_0 h^4(X_1)$ then $E_n|h^k(X_1)| \to E_0|h^k(X_1)|$ for $k \in \{1, 2, 3\}$ by applying Pratt's extended dominated convergence theorem from Appendix 2B [23].

In the sequel, let the operators $\mathrm{E}_n(\cdot)$ and $\mathrm{Var}_n(\cdot)$ denote expectation and variance with respect to a density that varies with (α_n, β_n).

Lemma 2.9. *Under the convergence conditions listed in Assumption 2.2, $\boldsymbol{Y}_n$ converges in distribution to a multivariate Gaussian distribution $\boldsymbol{Y}$:*

$$\begin{aligned} \boldsymbol{Y}_n = (Y_{1n}, Y_{2n}, Y_{3n}, Y_{4n})' &\overset{d(\beta_n)}{\rightarrow} \boldsymbol{Y} = (Y_1, Y_2, Y_3, Y_4)' \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_0) \\ \boldsymbol{\Sigma}_n \equiv \mathbf{Var}_n(\boldsymbol{Y}_n) &= \mathbf{Var}_n(\boldsymbol{Y}_1) + \mathbf{Var}_n(\boldsymbol{Y}_2) \overset{\beta_n}{\rightarrow} \boldsymbol{\Sigma}_0 \ . \end{aligned}$$

where $\boldsymbol{Y}_n, \boldsymbol{Y}_1, \boldsymbol{Y}_2$ are defined in (40) with (α_0, β_0) replaced by (α_n, β_n) such that at time index n

$$\begin{aligned} \boldsymbol{Y}_1 &\sim (\mathbf{E}_n(\boldsymbol{Y}_1), \mathbf{Var}_n(\boldsymbol{Y}_1)) \\ \boldsymbol{Y}_2 &\sim (\mathbf{E}_n(\boldsymbol{Y}_2), \mathbf{Var}_n(\boldsymbol{Y}_2)) \ . \end{aligned}$$

Proof: As shown in Lemma 2.8, the multivariate central limit theorem ([23], 2c.5) is applied to show the convergence in joint distribution of $\boldsymbol{Y}_n$

$$\begin{aligned} z_n = \boldsymbol{\lambda}' \boldsymbol{Y}_n &\overset{d(\beta_n)}{\rightarrow} z = \boldsymbol{\lambda}' \boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{\lambda}' \boldsymbol{\Sigma}_0 \boldsymbol{\lambda}\right) \\ \boldsymbol{\lambda} &= (\lambda_1, \lambda_2, \lambda_3, \lambda_4)' \ . \end{aligned}$$

The Lindeberg-Feller form of the central limit theorem ([30], Proposition 2.27) is applied to show the previous display. Let z_{ji}, C_n, and μ_{h^k} remain defined as in Lemma 2.8 such that for $i = 1, \dots, n_j$, $j = 1, 2$, and $k = 1, 2$

$$
\begin{aligned}
z_{ji} &\sim G_{n,z_{ji}} = G_{n,Z_j} \\
Z_j &\sim \left(\mathrm{E}_n\left(Z_j\right), \mathrm{Var}_n\left(Z_j\right)\right) = \left(0, \frac{1}{\rho_j}\boldsymbol{\lambda}'\mathbf{Var}_n(\boldsymbol{Y}_j)\boldsymbol{\lambda}\right) \\
C_n^2 &\equiv \sum_{i=1}^{n_1} \mathrm{Var}_n\left(z_{1i}\right) + \sum_{i=1}^{n_2} \mathrm{Var}_n\left(z_{2i}\right) \\
&= n_2 \boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda} \\
\mu_{h^k} &\equiv \mathrm{E}\left(h^k\left(X_2\right)\right) .
\end{aligned}
$$

As described in Remark 2.3, μ_{h^k} for $k = 1, 2$ do not depend on (α_n, β_n) so that the centering constants in the definition of $\boldsymbol{Y}_n$ in (40) do not vary with (α_n, β_n). The Lindeberg-Feller convergence condition, as specialized to z_{ji}/C_n, is satisfied for any $\varepsilon > 0$

$$
\begin{aligned}
&\left(\sum_{i=1}^{n_1} \int_{\left|\frac{z}{C_n}\right|>\varepsilon} \left(\frac{z}{C_n}\right)^2 dG_{n,z_{1i}}(z) + \sum_{i=1}^{n_2} \int_{\left|\frac{z}{C_n}\right|>\varepsilon} \left(\frac{z}{C_n}\right)^2 dG_{n,z_{2i}}(z)\right) \\
=&\frac{\rho_1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}} \int I\left(|z| > \varepsilon C_n\right) z^2 dG_{n,Z_1}(z) + \frac{1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}} \int I\left(|z| > \varepsilon C_n\right) z^2 dG_{n,Z_2}(z) \\
&\leq \frac{\sqrt{\rho_1} + 1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}} \int I\left(q(x|\boldsymbol{\lambda}) > \varepsilon C_n\right) q^2(x|\boldsymbol{\lambda}) g_2(x) dx \\
&\rightarrow 0 \text{ as } n \uparrow \infty
\end{aligned}
$$

where

$$
\begin{aligned}
q(x|\boldsymbol{\lambda}) &\equiv |\lambda_3| + \left(|\lambda_1| + |\lambda_4|\right)|h(x)| + |\lambda_2| h^2(x) \\
\sum_{i=1}^{n_1} &\mathrm{Var}\ \frac{z_{1i}}{C_n} + \sum_{i=1}^{n_2} \mathrm{Var}\ \frac{z_{2i}}{C_n} = 1
\end{aligned}
$$

since $\mathrm{Var}_n(z_n) = \boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda} \rightarrow \boldsymbol{\lambda}'\boldsymbol{\Sigma}_0\boldsymbol{\lambda} = \mathrm{Var}_0(z)$ and the integral converges to zero by applying Pratt's extended dominated convergence theorem from Appendix 2B [23], hence

$$
\frac{\sum_{i=1}^{n_1} z_{1i} + \sum_{i=1}^{n_2} z_{2i}}{\sqrt{n_2 \boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}}} \overset{d(\beta_n)}{\rightarrow} \mathrm{N}\left(0, 1\right)
$$

which proves the result that

$$
\boldsymbol{\lambda}'\boldsymbol{Y}_n \overset{d(\beta_n)}{\rightarrow} \mathrm{N}\left(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}_0\boldsymbol{\lambda}\right). \blacksquare
$$

In order to show that $\boldsymbol{D}_n \stackrel{P(\beta_n)}{\rightarrow} \boldsymbol{D}(\alpha_0, \beta_0)$ as $\beta_n \rightarrow \beta_0$, it suffices to prove the convergence results of (12), (13), (14), (15), and (16) as $\beta_n \rightarrow \beta_0$. These convergence results will be shown by proving uniform convergence results for the appropriate classes of functions using a specialized weak version of the abstract Glivenko-Cantelli Theorem.

Lemma 2.10. *Let X denote a random variable with a density function $p(x)$ and let $f(x)$ denote an integrable function with respect to $p(x)$ such that $\mu_f \equiv E(f(X)) < \infty$, then the characteristic function $\phi(t)$ of $f(X)$ is differentiable everywhere such that*

$$\begin{aligned} \frac{\phi(t+h)-\phi(t)}{h} &= E\Big(if(x)\,e^{itf(x)}f_h(x)\Big) \equiv \phi^h(t)\,,\ |f_h(x)| \le \sqrt{2} \\ \phi'(t) &\equiv \lim_{h\to 0}\phi^h(t) = E\Big(if(x)\,e^{itf(x)}\Big) \\ \phi'(0) &= i\mu_f\ . \end{aligned}$$

Proof: Direct calculation shows that the characteristic function $\phi(t)$ satisfies the following

$$\frac{\phi(t+h)-\phi(t)}{h} = \mathrm{E}\left(e^{itf(x)}\frac{(\cos(hf(x))-1)+i\sin(hf(x))}{h}\right) \equiv \phi^h(t)\ .$$

First order Taylor series expansions of $\cos(hf(x))$ and $\sin(hf(x))$ around $h=0$ shows

$$\begin{aligned} \cos(hf(x)) &= 1 - f(x)\sin(h_c f(x))\,h,\ h_c \in (0,h) \\ \sin(hf(x)) &= f(x)\cos(h_s f(x))\,h,\ h_s \in (0,h)\ . \end{aligned}$$

Hence the approximate derivative $\phi^h(t)$ of the characteristic function $\phi(t)$ can be rewritten as

$$\begin{aligned} \phi^h(t) &= \mathrm{E}\Big(if(x)\,e^{itf(x)}f_h(x)\Big) \\ f_h(x) &= \cos(h_s f(x)) + i\sin(h_c f(x))\,,\ h_c, h_s \in (0,h)\ . \end{aligned}$$

It is easy to see that for any fixed x

$$f_0(x) = 1,\ |f_h(x)|^2 = \cos^2(h_s f(x)) + \sin^2(h_c f(x)) \le 2\ .$$

Application of the dominated convergence theorem to $\phi^h(t)$ as $h \rightarrow 0$ under the assumption that the random variable $f(X)$ is integrable proves the final two results since

$$\left|\phi^h(t)\right| \le \mathrm{E}\left|if(x)\,e^{itf(x)}f_h(x)\right| \le \mathrm{E}\,|f(x)|\sqrt{2} < \infty\ . \blacksquare$$

Lemma 2.11. *Let $\{X_n : n = 1, 2, \dots\}$, denote a sequence of random variables with densities $p_n(x)$, and let X_0 denote another random variable with density $p_0(x)$ such that $p_n \to p_0$ almost everywhere. Let $f(x)$ denote a function that is integrable with respect to the sequence of densities $\{p_0, p_1, \dots\}$. If $E|f(X_n)| \to E|f(X_0)| < \infty$ then the sequence of characteristic functions $\phi_n(t)$ for $f(X_n)$ and the sequence of approximate derivatives $\phi_n^h(t)$ for the characteristic functions $\phi_n(t)$ converge uniformly to the characteristic function $\phi_0(t)$ for $f(X_0)$ and its approximate derivative $\phi_0^h(t)$*

$$\sup_t |\phi_n(t) - \phi_0(t)| \to 0$$

$$\sup_t \left|\phi_n^h(t) - \phi_0^h(t)\right| \to 0$$

$$\sup_t \left|\phi_n^t(0) - \phi_0^t(0)\right| \to 0 \ .$$

Proof: For the first result, applying Scheffe's convergence theorem involving densities (theorem XV) from [23] or applying Pratt's extended dominated convergence theorem from Appendix 2B [23], as $n \to \infty$ shows

$$\int |p_n(x) - p_0(x)|\, dx \to 0$$

since the integrand is dominated by $p_n(x)+p_0(x)$ such that as $p_n(x) \to p_0(x)$ almost everywhere and

$$\int (p_n(x) + p_0(x))\, dx = 2 \to 2 = 2\int p_0(x)\, dx < \infty \ .$$

For any t, the absolute difference between the characteristic functions is bounded by

$$|\phi_n(t) - \phi_0(t)| = \left|\int e^{itf(x)} (p_n(x) - p_0(x))\, dx\right| \leq \int |p_n(x) - p_0(x)|\, dx \ .$$

The three previous displays prove the first result that the the sequence of characteristic functions $\phi_n(t)$ of $f(X_n)$ converges uniformly to the characteristic function $\phi_0(t)$ of $f(X_0)$.

For the remaining results, assume without loss of generality that $\mathrm{E}|f(X_n)| < \infty$ for all n. Lemma 2.10 is applied to find a bound for the absolute difference between the approximate derivatives of the characteristic functions

$$\begin{aligned}\left|\phi_n^h(t) - \phi_0^h(t)\right| &= \left|\int if(x)\, e^{itf(x)} f_h(x) (p_n(x) - p_0(x))\, dx\right| \\ &\leq \int |f(x)| \sqrt{2}\, |p_n(x) - p_0(x)|\, dx \ .\end{aligned}$$

The integrand in the bound of the previous display is bounded by $|f(x)|\sqrt{2}(p_n(x)+p_0(x))$ such that as $p_n(x) \to p_0(x)$ almost everywhere and

$$\int |f(x)| \sqrt{2} (p_n(x) + p_0(x))\, dx \to 2\sqrt{2}\mathrm{E}\, |f(X_0)| < \infty .$$

Hence the remaining uniform convergence results for the sequence of approximate derivatives of the characteristic functions $\phi_n^h(t)$ and $\phi_n^t(0)$ for X_n is proven by applying Pratt's extended dominated convergence theorem from Appendix 2B [23]. ■

The following version of the weak law of large numbers is an extension of Proposition 2.16 in [30] to cover the case where the random sample densities p_n converge to a density p_0 almost everywhere.

Proposition 2.1. *Let $\{X_n : n = 1, 2, \dots\}$ denote a sequence of random variables with density functions $p_n(x)$ and let X_0 denote a random variable with density function $p_0(x)$ such that $p_n \to p_0$ almost everywhere. Let $f(x)$ denote an integrable function with respect to the sequence of densities $\{p_0, p_1, \dots\}$. Let $\rho > 0$ define a sample proportion and let $n_\rho \equiv n\rho/(1+\rho)$ define a sample size proportional to n. Let $\boldsymbol{x}_{n,\rho} \equiv \{x_{ni} : i = 1, \dots, n_\rho\}$ denote a random sample of size n_ρ from X_n, $n \in \{1, 2, \dots\}$. If $E|f(X_n)| \to E|f(X_0)|$ then*

$$\mathbb{P}_{n,\rho} f \equiv \frac{1}{n_\rho} \sum_{i=1}^{n_\rho} f(x_{ni}) \xrightarrow{P_n} P_0 f \equiv Ef(X_0) .$$

Proof: Let $\phi_n(t)$ denote the characteristic functions of $f(X_n)$ and let $\phi_0(t)$ denote the characteristic function of $f(X_0)$. By Lemma 2.10 the characteristic functions $\phi_n(t)$ for $f(X_n)$, $n \in \{0, 1, 2, \dots\}$ are differentiable for all t such that

$$\phi_n(t) = 1 + t\phi_n^t(0) .$$

Let $t_{n_\rho} \equiv t/n_\rho$. Applying Fubini's theorem shows for each fixed t that

$$\begin{aligned} \mathrm{E}e^{it\mathbb{P}_{n,\rho} f} &= \left(\phi_n\left(t_{n_\rho}\right)\right)^{n_\rho} = \left(\frac{\phi_n\left(t_{n_\rho}\right)}{\phi_0\left(t_{n_\rho}\right)}\right)^{n_\rho} \left(\phi_0\left(t_{n_\rho}\right)\right)^{n_\rho} \\ &= \left(1 + \frac{t}{n_\rho} \frac{\phi_n^{t_{n_\rho}}(0) - \phi_0^{t_{n_\rho}}(0)}{\phi_0\left(t_{n_\rho}\right)}\right)^{n_\rho} \left(1 + \frac{t}{n_\rho}\phi_0^{t_{n_\rho}}(0)\right)^{n_\rho} . \end{aligned}$$

By Lemma 2.11, the sequence of approximate derivatives $\phi_n^t(0)$ of the characteristic functions of $f(X_n)$ converges uniformly to the approximate derivative $\phi_0^t(0)$ of the characteristic function of $f(X_0)$, which shows as $n \to \infty$

$$\left|\frac{\phi_n^{t_{n_\rho}}(0) - \phi_0^{t_{n_\rho}}(0)}{\phi_0\left(t_{n_\rho}\right)}\right| = \frac{\left|\phi_n^{t_{n_\rho}}(0) - \phi_0^{t_{n_\rho}}(0)\right|}{\left|\phi_0\left(t_{n_\rho}\right)\right|} \leq \frac{\sup_{t^*}\left|\phi_n^{t^*}(0) - \phi_0^{t^*}(0)\right|}{\left|\phi_0\left(t_{n_\rho}\right)\right|}$$
$$\to \frac{0}{1} = 0 \, .$$

Lemma 2.10 also shows that $\phi_0^t(0)$ is continuous at $t = 0$ such that

$$\phi_0^{t_{n_\rho}}(0) \to \phi_0'(0) = i\mathrm{E}f\left(X_0\right) \text{ as } n_\rho \to \infty \, .$$

Combining the three previous displays shows the characteristic function for $\mathbb{P}_{n,\rho}f$ converges as $n \to \infty$

$$\mathrm{E}e^{it\mathbb{P}_{n,\rho}f} \to e^0 e^{t\phi_0'(0)} = e^{it\mathrm{E}f(X_0)} \, .$$

The previous display demonstrates pointwise convergence of the characteristic function for $\mathbb{P}_{n,\rho}f$ to the characteristic function of the constant random variable $\mathrm{E}f(X_0)$. By Levy's continuity theorem (Theorem 2.13 [30]), $\mathbb{P}_{n,\rho}f$ converges in distribution to $\mathrm{E}f(X_0)$. The result is proven since convergence in distribution to a constant implies convergence in probability. ■

Petrov (1995) [22] develops a weak law of large numbers result for triangular arrays of random variables. Under the assumptions of Proposition 2.1 the weak law of large numbers result of Theorem 4.11 [22] is valid if the following condition is met as $n \to \infty$ where m_n denotes the median of $p_n(x)$

$$n_1 \int \frac{\left((x - m_n)/n_1\right)^2}{1 + \left((x - m_n)/n_1\right)^2} p_n(x)\, dx \to 0 \, .$$

Pratt's extended dominated convergence theorem from Appendix 2B [23] is applied to show the previous convergence condition as $n \to \infty$ since

$$n_1 \int \frac{\left((x - m_n)/n_1\right)^2}{1 + \left((x - m_n)/n_1\right)^2} p_n(x)\, dx = \int \frac{\left(x - m_n\right)^2/n_1}{1 + \left((x - m_n)/n_1\right)^2} p_n(x)\, dx$$
$$|x - m_n| \leq |x| + |m_n| \, , \; \left|\frac{\left((x - m_n)/n_1\right)}{1 + \left((x - m_n)/n_1\right)^2}\right| \leq 1$$

since the integrand on the right hand side of the previous display converges pointwise to zero and since $\mathrm{E}_n(|X_1| + |m_n|) \to \mathrm{E}_0(|X_1| + |m_\infty|)$ under the assumption that each density in the sequence of densities $\{p_0, p_1, p_2, \dots\}$ has a unique median so that $m_\infty = m_0$ or under the assumption that the sequence of medians converges to a finite limit.

Theorem 2.3. *Let $\{X_n : n = 1, 2, \dots\}$ define a sequence of random variables with density functions $p_n(x)$ and let X_0 define a random variable with density function $p_0(x)$ such that $p_n \to p_0$ almost everywhere. Let $\mathcal{F} = \{f_{\boldsymbol{\theta}}(x) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ denote a parametric class of measurable functions and let $m(x)$ denote a measurable function as defined by Example 2.1 that are integrable with respect to the probability distributions $\{P_0, P_1, P_2, \dots\}$. Let $\rho > 0$ define a sample proportion and let $n_\rho \equiv n\rho/(1+\rho)$ define a sample size proportional to n. At time index n let $\boldsymbol{x}_{n,\rho} \equiv \{x_{ni} : i = 1, \dots, n_\rho\}$ denote a random sample from P_n. If $P_n|f_\theta| \equiv E|f_\theta(X_n)| \to P_0|f_\theta| \equiv E|f_\theta(X_0)|$ for all $f_\theta \in \mathcal{F}$ and $P_n m \equiv Em(X_n) \to P_0 m \equiv Em(X_0)$ as $n \to \infty$ then*

$$\|\mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}} - P_0 f_{\boldsymbol{\theta}}\|_{\mathcal{F}} \equiv \sup_{f_{\boldsymbol{\theta}} \in \mathcal{F}} \left| \frac{1}{n_\rho} \sum_{i=1}^{n_\rho} f_{\boldsymbol{\theta}}(x_{ni}) - E f_{\boldsymbol{\theta}}(X_0) \right| \xrightarrow{P_n} 0 \ .$$

Proof: Given a bracket size of ϵ, Example 2.1 implies that in order to cover $\mathcal{F}$ with a finite number of ϵ-brackets in $L_1(P_0)$ it is sufficient to cover $\boldsymbol{\Theta}$ with a finite number of balls of diameter $\epsilon/(2P_0 m)$. Example 2.1 bounds the minimum number of ϵ-brackets in $L_1(P_0)$ needed to cover $\mathcal{F}$ by

$$N_{[]}(\epsilon, \mathcal{F}, L_1(P_0)) \le K \left(\frac{\text{diam } \boldsymbol{\Theta} \times P_0 m}{\epsilon} \right)^d \ .$$

Let $N_{\mathcal{F},\epsilon} \equiv N_{[]}(\epsilon, \mathcal{F}, L_1(P_0))$ and let $\mathcal{F}_{\epsilon,j} \equiv \{f_{\boldsymbol{\theta}} \in \mathcal{F} : f_{\boldsymbol{\theta}} \in j\text{th } \epsilon\text{-bracket}\}$, for $j = 1, \dots, N_{\mathcal{F},\epsilon}$. Choose a single function in each parametric subclass $f_{\boldsymbol{\theta}} \in \mathcal{F}_{\epsilon,j}$ and denote it as $f_{\boldsymbol{\theta}_{(j)}}$ for $j = 1, \dots, N_{\mathcal{F},\epsilon}$ and let $\mathcal{F}_\epsilon \equiv \{f_{\boldsymbol{\theta}_{(j)}} : j = 1, \dots, N_{\mathcal{F},\epsilon}\}$. The j-th ϵ-bracket of the form $[l_j, u_j]$ is constructed using $f_{\boldsymbol{\theta}_{(j)}}$ such that for $f_{\boldsymbol{\theta}_{(j)}}, f_{\boldsymbol{\theta}} \in \mathcal{F}_{\epsilon,j}$ where $\|\boldsymbol{\theta}_{(j)} - \boldsymbol{\theta}\| \le \epsilon/(2P_0 m)$

$$l_j \equiv f_{\boldsymbol{\theta}_{(j)}} - \frac{\epsilon}{2P_0 m} m \le f_{\boldsymbol{\theta}} \le f_{\boldsymbol{\theta}_{(j)}} + \frac{\epsilon}{2P_0 m} m \equiv u_j, \ P_0(u_j - l_j) = \epsilon \ .$$

For any $f_{\boldsymbol{\theta}} \in \mathcal{F}$ and a bracket size ϵ there exist an $\mathcal{F}_{\epsilon,j}$ with $f_{\boldsymbol{\theta}}, f_{\boldsymbol{\theta}_{(j)}} \in \mathcal{F}_{\epsilon,j}$. Applying the ϵ-bracket inequalities from the previous display shows that

$$|\mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}} - P_0 f_{\boldsymbol{\theta}}| \le \left| \mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}_{(j)}} - P_0 f_{\boldsymbol{\theta}_{(j)}} \right| + \frac{\epsilon}{2P_0 m} |\mathbb{P}_{n,\rho} m - P_0 m| + \epsilon \ .$$

The previous display, true for any $f_{\boldsymbol{\theta}} \in \mathcal{F}_{\epsilon,j}$ given an ϵ-bracket, implies the following supremum over all $f_{\boldsymbol{\theta}} \in \mathcal{F}$ given an ϵ-bracket

$$\sup_{f_{\boldsymbol{\theta}} \in \mathcal{F}} |\mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}} - P_0 f_{\boldsymbol{\theta}}| \le \sup_{f_{\boldsymbol{\theta}_{(j)}} \in \mathcal{F}_\epsilon} \left| \mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}_{(j)}} - P_0 f_{\boldsymbol{\theta}_{(j)}} \right| + \frac{\epsilon}{2P_0 m} |\mathbb{P}_{n,\rho} m - P_0 m| + \epsilon \ .$$

Given $\eta, \varepsilon > 0$ choose a bracket size $\epsilon \leq \eta/3$ and choose $N_{\eta,\varepsilon}$ by applying the weak law of large numbers from Proposition 2.1 such that for $n > N_{\eta,\varepsilon}$

$$P\left(\left|\mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}_{(j)}} - P_0 f_{\boldsymbol{\theta}_{(j)}}\right| < \frac{\eta}{3}\right) > 1 - \frac{\varepsilon}{2N_{\mathcal{F},\epsilon}} \text{ for all } f_{\boldsymbol{\theta}_{(j)}} \in \mathcal{F}_\epsilon$$

$$P\left(\frac{\epsilon}{2P_0 m}\left|\mathbb{P}_{n,\rho} m - P_0 m\right| < \frac{\eta}{3}\right) \geq P\left(\left|\mathbb{P}_{n,\rho} m - P_0 m\right| < 2P_0 m\right) > 1 - \frac{\varepsilon}{2}\ .$$

Hence for $n > N_{\eta,\varepsilon}$ the previous two displays show that

$$P\left(\sup_{f_{\boldsymbol{\theta}_{(j)}} \in \mathcal{F}_\epsilon}\left|\mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}_{(j)}} - P_0 f_{\boldsymbol{\theta}_{(j)}}\right| < \frac{\eta}{3}\right) > 1 - \frac{\varepsilon}{2}$$

$$P\left(\sup_{f_{\boldsymbol{\theta}} \in \mathcal{F}}\left|\mathbb{P}_{n,\rho} f_{\boldsymbol{\theta}} - P_0 f_{\boldsymbol{\theta}}\right| < \eta\right) > 1 - \varepsilon\ .$$

The result is proven since $\eta, \varepsilon > 0$ are arbitrary. ■

The asymptotic properties of extremum estimators from Amemiya [1] is applied to show the convergence in probability property of the estimators $(\hat{\alpha}, \hat{\beta}) \to (\alpha_0, \beta_0)$ as $n \to \infty$. As defined previously prior to Lemma 2.4, let $l_n(\alpha, \beta) = l(\alpha, \beta) + n \log(n_2)$.

Lemma 2.12. *Under the first four convergence conditions of Assumption 2.2, if $h(x)$ is integrable with respect to the sequence of densities $\{g_2, p_0, p_1, \dots\}$ such that $E_n|h(X_1)| \to E_0|h(X_1)|$ and if $h^2(x)$ is integrable with respect to the densities $\{g_2, p_0\}$, then $(\hat{\alpha}, \hat{\beta})$ converges in probability to (α_0, β_0).*

Proof: Let $\boldsymbol{\Theta}$ define a bounded compact subspace of $\mathbb{R}^2$ that includes the sequence of distortion parameters (α_n, β_n) for $n \in \{1, 2, \dots\}$ and includes the limiting distortion parameters (α_0, β_0) such that $(\alpha_n, \beta_n) \to (\alpha_0, \beta_0)$. Let $\boldsymbol{\Theta}_n^* = \{(\alpha_*, \beta_*) : l_n(\alpha_*, \beta_*) = \max_{(\alpha,\beta)\in\boldsymbol{\Theta}} l_n(\alpha, \beta)\}$. Application of Theorem 4.1.1 from Amemiya [1], shows that $(\alpha_*, \beta_*) \in \boldsymbol{\Theta}_n^*$ converges in probability to (α_0, β_0) under the following conditions

- **(A)** The parameter subspace $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^2$ that includes (α_0, β_0),
- **(B)** $l_n(\alpha, \beta)$ is continuous in $(\alpha, \beta) \in \boldsymbol{\Theta}$ for all $\boldsymbol{t} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$ and is a measurable function of $\boldsymbol{t}$ for all $(\alpha, \beta) \in \boldsymbol{\Theta}$,
- **(C)** $l_n(\alpha, \beta)$ converges to a nonstochastic function $g(\alpha, \beta)$ in probability uniformly in $(\alpha, \beta) \in \boldsymbol{\Theta}$ as $n \to \infty$, and $g(\alpha, \beta)$ attains a unique global maximum at (α_0, β_0).

Condition (A) is satisfied by construction. Condition (B) is also satisfied since the profile log-likelihood equation $l(\alpha, \beta)$ is continuous in $(\alpha, \beta) \in \boldsymbol{\Theta}$ and since $h(x)$ is integrable with respect to the densities $g_1(x)$ and $g_2(x)$.

With regard to condition (C), Definition 2.6 defines two classes of functions $\mathcal{F}_1(\boldsymbol{\Theta})$ and $\mathcal{F}_2(\boldsymbol{\Theta})$ with parametric index $\boldsymbol{\Theta}$ that are used to prove the uniform convergence in probability condition (D) for Lemma 2.4. The proof of condition (D) for Lemma 2.4 shows that $\mathcal{F}_1(\boldsymbol{\Theta})$ and $\mathcal{F}_2(\boldsymbol{\Theta})$ are parametric classes with a common Lipschitz bound $m(x) \equiv 2(1 + |h(x)|)$. By assumption $h(x)$ is integrable with respect to the sequence of densities $\{g_2, p_0, p_1, \dots\}$ such that $\mathrm{E}_n|h(X_1)| \to \mathrm{E}_0|h(X_1)|$. Hence $m(x)$ and the functions $f_1(x|\alpha, \beta) \in \mathcal{F}_1(\boldsymbol{\Theta})$ are also integrable to the same sequence of densities such that $\mathrm{E}_n|m(X_1)| \to \mathrm{E}_0|m(X_1)|$ and $\mathrm{E}_n|f_1(X_1|\alpha, \beta)| \to \mathrm{E}_0|f_1(X_1|\alpha, \beta)|$ by applying Pratt's extended dominated convergence theorem from Appendix 2B [23]. The specialized weak Glivenko-Cantelli Theorem 2.3 is applied to show for $f_1(x|\alpha, \beta) \in \mathcal{F}_1(\boldsymbol{\Theta})$

$$\sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} f_1\left(x_{1i}|\alpha, \beta\right) - \mathrm{E}\left(f_1\left(X_1|\alpha, \beta\right)\right) \right| \stackrel{P_n}{\to} 0 \, .$$

Lemma 2.2 was previously applied to $f_2(x|\alpha, \beta) \in \mathcal{F}_2(\boldsymbol{\Theta})$ to show

$$\sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \left| \frac{1}{n_2} \sum_{i=1}^{n_2} f_2\left(x_{2i}|\alpha, \beta\right) - \mathrm{E}\left(f_2\left(X_2|\alpha, \beta\right)\right) \right| \stackrel{as*}{\to} 0 \, .$$

The combination of the two previous displays proves the uniform convergence in probability condition

$$\begin{aligned}
&\sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \left| \frac{1}{n} l_n\left(\alpha, \beta\right) - g\left(\alpha, \beta\right) \right| \\
\leq &\sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \frac{\rho_1}{1+\rho_1} \left| \frac{1}{n_1} \sum_{i=1}^{n_1} f_1\left(x_{1i}|\alpha, \beta\right) - \mathrm{E}\left(f_1\left(X_1|\alpha, \beta\right)\right) \right| \\
+ &\sup_{(\alpha,\beta)\in\boldsymbol{\Theta}} \frac{1}{1+\rho_1} \left| \frac{1}{n_2} \sum_{i=1}^{n_2} f_2\left(x_{2i}|\alpha, \beta\right) - \mathrm{E}\left(f_2\left(X_2|\alpha, \beta\right)\right) \right| \\
&\stackrel{P_n}{\to} 0 \, .
\end{aligned}$$

The function $g(\alpha, \beta)$ and its gradient and hessian have the following forms, as shown in the proof of Lemma 2.4 for condition (D), and as shown in the alternate proof of Lemma 2.5, under the limit condition that $X_1 \sim g_1 = p_0$

$$g(\alpha, \beta) = \mathrm{E}\frac{1}{n} l_n(\alpha, \beta)$$
$$\boldsymbol{\nabla} g(\alpha, \beta) = \mathrm{E}\frac{1}{n} \boldsymbol{\nabla} l_n(\alpha, \beta)$$
$$\boldsymbol{\nabla}\boldsymbol{\nabla}' g(\alpha, \beta) = \mathrm{E}\frac{1}{n} \boldsymbol{\nabla}\boldsymbol{\nabla}' l_n(\alpha, \beta) \ .$$

The actual form of $g(\alpha, \beta)$, its gradient $\boldsymbol{\nabla} g(\alpha, \beta)$, and its hessian $\boldsymbol{\nabla}\boldsymbol{\nabla}' g(\alpha, \beta)$, are identified in (37), (36), and (38) within Lemma 2.4. The proof of Lemma 2.5 shows that $g(\alpha, \beta)$ has a global maximum at (α_0, β_0) where $\boldsymbol{\nabla} g(\alpha_0, \beta_0) = \mathbf{0}$ and where the hessian $\boldsymbol{\nabla}\boldsymbol{\nabla}' g(\alpha, \beta)$ is positive definite for all $(\alpha, \beta) \in \mathbb{R}^2$ under the assumption that $h(x)$ is non-constant with respect to g_2. Hence the proof that $(\alpha_*, \beta_*) \in \boldsymbol{\Theta}_n^*$ converges in probability to (α_0, β_0) is complete.

Lemma 2.6 is applied to complete the proof that $(\hat{\alpha}, \hat{\beta})$ converges in probability to (α_0, β_0), since $n^{-1} l_n(\alpha, \beta)$ converges uniformly in probability to $g(\alpha, \beta)$ for $(\alpha, \beta) \in \boldsymbol{\Theta}$ where $g(\alpha, \beta)$ has a global maximum at (α_0, β_0), and since $h(x)$ is non-constant with respect to g_2 by assumption. Hence the result is proven. ■

The proofs of Corollaries 2.9 and 2.10 remain valid as follows.

Corollary 2.11. *Under the convergence conditions of Lemma 2.12, if* $(\hat{\alpha}, \hat{\beta}) \stackrel{P(\beta_n)}{\rightarrow} (\alpha_0, \beta_0)$ *then* $(\acute{\alpha}, \acute{\beta})$ *and* $(\grave{\alpha}, \grave{\beta}) \stackrel{P(\beta_n)}{\rightarrow} (\alpha_0, \beta_0)$. ■

The following Lemma 2.13 provides a counterpart to Lemma 2.3 as $(\alpha_n, \beta_n) \rightarrow (\alpha_0, \beta_0)$. This lemma utilizes the abstract parametric classes $\mathcal{F}_j(f_1, f_2)$ with parametric index $\boldsymbol{\Theta}$ for $j = 1, 2$ from Definition 2.3. This lemma, with $\mathcal{F}_j(f_1, f_2)$ for $j = 1, 2$ specialized to $\mathcal{F}_{j|k}^{(1)}(\boldsymbol{\Theta})$ for $k = 1, 2$ from Definition 2.4 and specialized to $\mathcal{F}_{j|k}^{(2)}(\boldsymbol{\Theta})$ for $k = 0, 1, 2, 3$ from Definition 2.4, is applied in Lemma 2.14 to show specific random sample convergence results as $(\alpha_n, \beta_n) \rightarrow (\alpha_0, \beta_0)$.

Lemma 2.13. *Under the first three convergence conditions of Assumption 2.2 with* $m = 2$ *and with the parametric classes of functions* $\mathcal{F}_j(f_1, f_2)$ *with parametric index* $\boldsymbol{\Theta}$ *for* $j = 1, 2$ *from Definition 2.3 with Lipschitz bounds* $m_j(x)$*, if the functions* $f(x|\alpha, \beta) \in \mathcal{F}_1(f_1, f_2)$ *and* $m_1(x)$ *are integrable with respect to the sequence of densities* $\{p_0, p_1, p_2, \dots\}$ *such that* $E_n|f(X_1|\alpha, \beta)| \rightarrow E_0|f(X_1|\alpha, \beta)|$ *and* $E_n m_1(X_1) \rightarrow E_0 m_1(X_1)$*, if the functions* $f \in \mathcal{F}_2(f_1, f_2)$ *and* $m_2(x)$ *are integrable with respect to the density* g_2*,* *and if* $(\alpha_*, \beta_*) \stackrel{P(\beta_n)}{\rightarrow} (\alpha_0, \beta_0) \in \mathbf{\Theta}$*, then*

$$\sum_{i=1}^{n} f_1(t_i)\, f_2(t_i|\alpha_n, \beta_n)\, \hat{p}(\alpha_n, \beta_n) \stackrel{P(\beta_n)}{\rightarrow} E(f_1(X_2)\, f_2(X_2|\alpha_0, \beta_0))$$

$$\sum_{i=1}^{n} f_1(t_i)\, f_2(t_i|\alpha_*, \beta_*)\, \hat{p}(\alpha_*, \beta_*) \stackrel{P(\beta_n)}{\rightarrow} E(f_1(X_2)\, f_2(X_2|\alpha_0, \beta_0)) \ .$$

Proof: The proof of this lemma makes use of expressions (23) and (24) from Lemma 2.3. Expression (23) for $j = 1$ converges in probability to 0 for any nonrandom sequence $(\alpha_*, \beta_*) = (\alpha_n, \beta_n) \to (\alpha_0, \beta_0)$ by applying Theorem 2.3 with $\mathcal{F}$ specialized to $\mathcal{F}_1(f_1, f_2)$ with parametric index $\mathbf{\Theta}$ and by applying Proposition 2.1 with $f(x)$ specialized to $m_1(x)$. Expression (23) for $j = 2$ converges almost surely to 0 for any nonrandom sequence $(\alpha_*, \beta_*) = (\alpha_n, \beta_n) \to (\alpha_0, \beta_0)$ by applying Lemma 2.2 and by applying the strong law of large numbers. Combining in (24) the convergence results from the two previous statements for $j \in \{1, 2\}$ proves the first result.

Expression (23) for $j \in \{1, 2\}$ converges in probability to 0 for $(\alpha_*, \beta_*) \stackrel{P(\beta_n)}{\rightarrow} (\alpha_0, \beta_0)$ by applying Theorem 2.3, Proposition 2.1, Lemma 2.2, the weak law of large numbers, and by applying Slutsky's theorem. Combining in (24) the two convergence results from the previous statement for $j \in \{1, 2\}$ proves the second result. ■

Lemma 2.14. *Under the convergence conditions defined in Assumption 2.2*

$$\hat{\mu}(\alpha_n, \beta_n) \stackrel{P(\beta_n)}{\rightarrow} \mu_h$$

$$\hat{\sigma}_h^2\left(\hat{\alpha}, \hat{\beta}\right) \stackrel{P(\beta_n)}{\rightarrow} \sigma_h^2$$

$$\nabla \hat{\sigma}_h^2(\alpha, \beta)\big|_{(\acute{\alpha}, \acute{\beta})} \stackrel{P(\beta_n)}{\rightarrow} \nabla \sigma_h^2(\alpha_0, \beta_0)$$

$$-\frac{1}{n}\, \nabla\nabla' l(\alpha, \beta)\big|_{(\grave{\alpha}, \grave{\beta})} \stackrel{P(\beta_n)}{\rightarrow} \mathbf{S}(\alpha_0, \beta_0) \ .$$

Proof: Let $\mathbf{\Theta}$ denote a bounded subset of $\mathbb{R}^2$ that contains $\{(\alpha_n, \beta_n) : n = 0, 1, 2, \dots\}$. As previously shown, the functions $f \in \mathcal{F}_{j|k}^{(1)}(\mathbf{\Theta}) \equiv \mathcal{F}_j(h^k(x), 1)$, $j \in \{1, 2\}$ with parametric index $\mathbf{\Theta}$, $k \in \{1, 2\}$, have Lipschitz bounds $m_{j|k}^{(1)}(x) \equiv \rho_j(|h^k(x)| + |k^{k+1}(x)|)$. Also the functions $f \in \mathcal{F}_{j|k}^{(2)}(\mathbf{\Theta}) \equiv \mathcal{F}_j(h^k(x), \rho_1 w_1(x|\alpha, \beta)/D_1(x|\alpha, \beta))$ with parametric index $\mathbf{\Theta}$, $j \in \{1, 2\}$, $k \in \{0, 1, 2, 3\}$, have Lipschitz bounds $m_{j|k}^{(2)}(x) \equiv 3\rho_j(|h^k(x)| + |k^{k+1}(x)|)$. Under the assumptions of this lemma, the functions $f \in \mathcal{F}_{1|k}^{(1)}(\mathbf{\Theta})$ and

$m^{(1)}_{1|k}(x)$ are integrable with respect to the sequence of densities $\{p_0, p_1, \dots\}$ for $k \in \{1,2\}$. The functions $f \in \mathcal{F}^{(1)}_{1|k}(\mathbf{\Theta})$ are bounded by $|h^k(x)|$ such that $P_n|f| \to P_0|f|$ by applying Pratt's extended dominated convergence theorem from Appendix 2B [23]. The Lipschitz bounds $m^{(1)}_{1|k}(x)$ converge under the X_1 densities such that $P_n m^{(1)}_{1|k} \to P_0 m^{(1)}_{1|k}$. Also the functions $f \in \mathcal{F}^{(1)}_{2|k}(\mathbf{\Theta})$ and $m^{(1)}_{2|k}(x)$ are integrable with respect to the density g_2. Similarly, the functions $f \in \mathcal{F}^{(2)}_{1|k}(\mathbf{\Theta})$ and $m^{(2)}_{1|k}(x)$ are integrable with respect to the sequence of densities $\{p_0, p_1, \dots\}$ for $k \in \{0,1,2,3\}$. The functions $f \in \mathcal{F}^{(2)}_{1|k}(\mathbf{\Theta})$ are bounded by $|h^k(x)|$ such that $P_n|f| \to P_0|f|$ by applying Pratt's extended dominated convergence theorem from Appendix 2B [23]. The Lipschitz bounds $m^{(2)}_{1|k}(x)$ converge under the X_1 densities such that $P_n m^{(2)}_{1|k} \to P_0 m^{(2)}_{1|k}$. Also the functions $f \in \mathcal{F}^{(2)}_{2|k}(\mathbf{\Theta})$ and $m^{(2)}_{2|k}(x)$ are integrable with respect to the density g_2.

Lemma 2.12 shows that $(\hat{\alpha}, \hat{\beta}) \overset{P(\beta_n)}{\to} (\alpha_0, \beta_0)$ under the assumptions of this lemma.

The first result is proven, under the assumptions of this lemma, by starting with the definition of $\hat{\mu}_{h^k}$ from (26), and applying Lemma 2.13 to the functions $f \in \mathcal{F}^{(1)}_{j|k}(\mathbf{\Theta})$ for $j \in \{1,2\}$ and $k = 1$ as $(\alpha_n, \beta_n) \to (\alpha_0, \beta_0)$.

The second result is proven, under the assumptions of this lemma, by starting with the definition of $\hat{\sigma}^2_h$ from (28), and applying Lemma 2.13 to the functions $f \in \mathcal{F}^{(1)}_{j|k}(\mathbf{\Theta})$ for $j \in \{1,2\}$ and $k \in \{1,2\}$ as $(\hat{\alpha}, \hat{\beta}) \overset{P(\beta_n)}{\to} (\alpha_0, \beta_0)$.

The third result is proven, under the assumptions of this lemma, by starting with (29) and (30), applying Corollary 2.11 to $(\acute{\alpha}, \acute{\beta})$, applying Lemma 2.13 to the functions $f \in \mathcal{F}^{(1)}_{j|k}(\mathbf{\Theta})$ for $j \in \{1,2\}$ and $k = 1$ as $(\acute{\alpha}, \acute{\beta}) \overset{P(\beta_n)}{\to} (\alpha_0, \beta_0)$, applying Lemma 2.13 to the functions $f \in \mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta})$ for $j \in \{1,2\}$ and $k \in \{1,2,3\}$ as $(\acute{\alpha}, \acute{\beta}) \overset{P(\beta_n)}{\to} (\alpha_0, \beta_0)$, and by applying Slutsky's theorem.

The fourth result is proven, under the assumptions of this lemma, by starting with (33), applying Corollary 2.11 to $(\grave{\alpha}, \grave{\beta})$, applying Lemma 2.13 to the functions $f \in \mathcal{F}^{(2)}_{j|k}(\mathbf{\Theta})$ for $j \in \{1,2\}$ and $k \in \{0,1,2\}$ as $(\grave{\alpha}, \grave{\beta}) \overset{P(\beta_n)}{\to} (\alpha_0, \beta_0)$, and applying Slutsky's theorem. ■

Lemma 2.15. *Under the convergence conditions defined in Assumption 2.2,*

$\boldsymbol{D}_n$ *converges in probability to* $\boldsymbol{D}(\alpha_0, \beta_0)$ *as defined in (39)*

$$\boldsymbol{D}_n \overset{P(\beta_n)}{\rightarrow} \boldsymbol{D}(\alpha_0, \beta_0) = \frac{1}{2\sigma_h} \left(-2\mu_h \beta_0, \quad \beta_0, \quad \mathbf{Q}(\alpha_0, \beta_0)\, \mathbf{S}^{-1}(\alpha_0, \beta_0)\right)' .$$

Proof: Lemma 2.14, the continuous mapping theorem, and Slutsky's theorem are applied to prove the result. ■

Theorem 2.4. *Under the convergence conditions of Assumption 2.2,* $\tilde{Z}_n^*$ *converges to a Gaussian random variable* $\tilde{Z}^*$.

Proof: The convergence in distribution of $\tilde{\mathrm{Z}}_n^*$ as $\beta_n \to \beta_0$ is established using Slutsky's theorem, Lemma 2.15, and Lemma 2.9

$$\tilde{\mathrm{Z}}_n^* = \boldsymbol{D}_n' \boldsymbol{Y}_n \overset{d(\beta_n)}{\rightarrow} \tilde{\mathrm{Z}}^* = \boldsymbol{D}' \boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{D}' \boldsymbol{\Sigma} \boldsymbol{D}\right)$$

where $\boldsymbol{D} = \boldsymbol{D}(\alpha_0, \beta_0)$ as defined in (39). ■

Corollary 2.12. *If the limiting distortion parameters* (α_0, β_0) *identify a null distortion* $(0, 0)$ *then the limiting distribution of* $\tilde{Z}^*$ *is a standard Gaussian distribution:* $\tilde{Z}^* \sim N(0, 1)$.

Proof: Direct calculations are used to show the following

$$\mathbf{Q}(0,0) = \left(0, 2\sigma_h^2\right), \; \mathbf{S}(0,0) = \frac{\rho_1}{(1+\rho_1)^2} \begin{bmatrix} 1 & \mu_h \\ \mu_h & \mu_{h^2} \end{bmatrix}, \; \mathbf{V}_0 = \frac{\rho_1^2 \sigma_h^2}{(1+\rho_1)^4} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} .$$

The result is proven by first calculating $\boldsymbol{D}(0, 0)$

$$\boldsymbol{D}(0,0) = \frac{(1+\rho_1)^2}{\rho_1 \sigma_h} \left(0, \quad 0, \quad -\mu_h, \quad 1\right)' = \left(D_1, \quad D_2, \quad D_3, \quad D_4\right)'$$

$$\boldsymbol{D}' \boldsymbol{\Sigma} \boldsymbol{D} = \left(D_3 \quad D_4\right) \mathbf{V}_0 \begin{pmatrix} D_3 \\ D_4 \end{pmatrix} = 1 . \; ■$$

2.2.1.1 Gaussian Example In this section, an example of the asymptotic $\tilde{\mathrm{Z}}^*$ distribution is calculated where X_1 and X_2 have Gaussian distributions with different means μ_1 and μ_2 and with a common variance $\sigma^2 = 1$, as described in section 2.1.1.1.

Concerning the convergence conditions of Assumption 2.1, $h(x)$ is continuous and non-constant with respect to the Gaussian density, and $h^k(x) = x^k$ is integrable with respect to the Gaussian density for $k = 1, 2, 3, 4$ as identified in section 2.1.1.1. Hence the convergence conditions are met that allow $(\hat{\alpha}, \hat{\beta})$ to converge in probability to their true value (α_0, β_0). Also the convergence results of (13), (14), (15), (16), (17) are valid. In conclusion, $\tilde{\mathrm{Z}}_n^*$ converges in distribution to $\tilde{\mathrm{Z}}^*$ identified in (44). Figure 1 graphs the variance of $\tilde{\mathrm{Z}}^*$ versus the difference in means of X_1 and X_2.

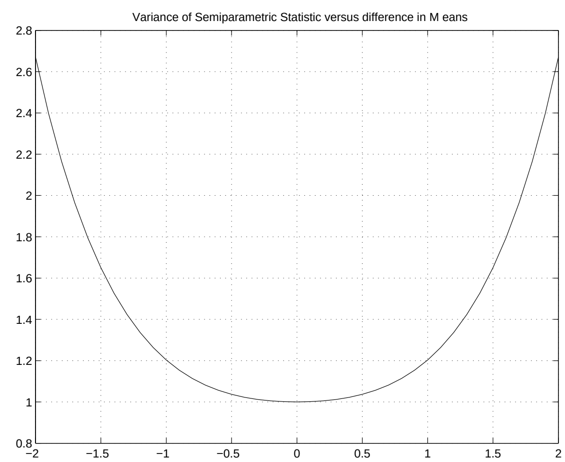

Figure 1: Variance of $\tilde{\mathrm{Z}}^*$ versus $\mu_1 - \mu_2$ when $X_1 \sim \mathrm{N}(\mu_1, 1)$, $X_2 \sim \mathrm{N}(\mu_2, 1)$.

Concerning the convergence conditions of Assumption 2.2, the Gaussian density $p(x|\mu, \sigma^2)$ is a continuous function of its parameters (μ, σ^2) such than $g_1(x) = p(x|\mu_1, \sigma^2) \to g_2(x) = p(x|\mu_2, \sigma^2)$ for all $x \in \mathbb{R}$ as $\mu_1 \to \mu_2$. The distortion parameters (α_n, β_n) are also continuous functions of the Gaussian parameters (μ, σ^2) as identified in section 2.1.1.1 such that $(\alpha_n, \beta_n) \to (0, 0)$ as $\mu_1 \to \mu_2$. The function $h(x) = x$ is continuous and non-constant with respect to the Gaussian density, $h^k(x) = x^k$ is integrable with respect to the sequence of Gaussian densities for $k \in \{1, 2, 3, 4\}$ as identified in section 2.1.1.1, and $E_n X_1^4 \to E_0 X_1^4$ as $\mu_1 \to \mu_2$. Hence the convergence conditions have been met that allow $\tilde{\mathrm{Z}}_n^*$ to converge in distribution to $\tilde{\mathrm{Z}}^* \sim \mathrm{N}(0, 1)$ as $\beta_n \to 0$.

2.2.1.2 Gamma Examples I and II In this section, two examples of the asymptotic $\tilde{Z}^*$ distribution are calculated using gamma distributions. For Example I, X_1 and X_2 have gamma distributions with a common shape parameter $\alpha_\gamma = 1$ and with different scale parameters $\beta_{\gamma 1}$ and $\beta_{\gamma 2}$ as described in section 2.1.1.2. For Example II, X_1 and X_2 have gamma distributions with different shape parameters $\alpha_{\gamma 1}$ and $\alpha_{\gamma 2}$ and with a common scale parameter $\beta_\gamma = 1$ as described in section 2.1.1.3.

Concerning the convergence conditions of Assumption 2.1 for the Gamma I example, $h(x)$ is continuous and non-constant with respect to the gamma density, and $h^k(x) = x^k$ is continuous, non-constant, and integrable with respect to the gamma density, for $k = 1, 2, 3, 4$ as identified in section 2.1.1.2. For the Gamma II example, $h(x)$ is continuous and non-constant with respect to the gamma density, and $h^k(x) = \log^k(x)$ is integrable with respect to the gamma density, for $k = 1, 2, 3, 4$, since the moment generating function, $\mathrm{M}_{\log(X_j)}(t)$ $j = 1, 2$, exists for t in a neighborhood of 0 as identified in section 2.1.1.3, see Cassela and Berger (1990) [6] Definition 2.3.3 and Theorem 2.3.2. Hence the conditions are met that allow $(\hat{\alpha}, \hat{\beta})$ to converge in probability to their true value (α_0, β_0). Also for both examples, the convergence results of (13), (14), (15), (16), (17) are valid. In conclusion, $\tilde{Z}^*_n$ converges in distribution to $\tilde{Z}^*$ identified in (44).

Concerning the convergence conditions of Assumption 2.2 for the Gamma I and II examples, the gamma density $p(x|\alpha_\gamma, \beta_\gamma)$ is a continuous function of its parameters $(\alpha_\gamma, \beta_\gamma)$ such that $g_1(x) = p(x|\alpha_\gamma, \beta_{\gamma 1}) \to g_2(x) = p(x|\alpha_\gamma, \beta_{\gamma 2})$ for all $x \in \mathbb{R}^+$ as $\beta_{\gamma 1} \to \beta_{\gamma 2}$ and $g_1(x) = p(x|\alpha_{\gamma 1}, \beta_\gamma) \to g_2(x) = p(x|\alpha_{\gamma 2}, \beta_\gamma)$ for all $x \in \mathbb{R}^+$ as $\alpha_{\gamma 1} \to \alpha_{\gamma 2}$. The distortion parameters (α_n, β_n) are also continuous functions of the gamma parameters $(\alpha_\gamma, \beta_\gamma)$ as identified in sections 2.1.1.2 and 2.1.1.3 such that $(\alpha_n, \beta_n) \to (0, 0)$ for both Gamma I and II examples as $\beta_{\gamma 1} \to \beta_{\gamma 2}$ or $\alpha_{\gamma 1} \to \alpha_{\gamma 2}$. The functions $h(x) = x$ for the Gamma I example, and $h(x) = \log(x)$ for the Gamma II example, are continuous and non-constant with respect to the gamma density; $h^k(x)$ is integrable with respect to the sequence of gamma densities for $k \in \{1, 2, 3, 4\}$ as identified in sections 2.1.1.2, 2.1.1.3, and above; $E_n X_1^4 \to E_0 X_1^4$ as $\beta_{\gamma 1} \to \beta_{\gamma 2}$; and $E_n \log^4(X_1) \to E_0 \log^4(X_1)$ as $\alpha_{\gamma 1} \to \alpha_{\gamma 2}$ since the moment generating functions converge. Hence the convergence conditions have been met that allow $\tilde{Z}^*_n$ to converge in distribution to $\tilde{Z}^* \sim \mathrm{N}(0, 1)$ as $\beta_n \to 0$.

Figure 2 graphs the variance of $\tilde{Z}^*$ versus a range of $\beta_{\gamma 1}$ parameter values for X_1 with $\beta_{\gamma 2} = 3$ for X_2. Figure 3 graphs the variance of $\tilde{Z}^*$ versus a range of $\alpha_{\gamma 1}$ parameter values for X_1 with $\alpha_{\gamma 2} = 3$ for X_2.

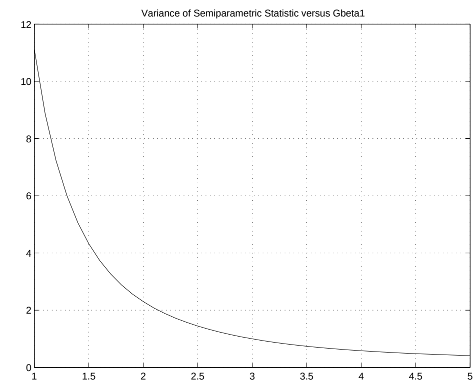

Figure 2: Variance of $\tilde{Z}^*$ versus $\beta_{\gamma 1}$ when $X_1 \sim \text{Gamma}(1, \beta_{\gamma 1})$, $X_2 \sim \text{Gamma}(1, 3)$.

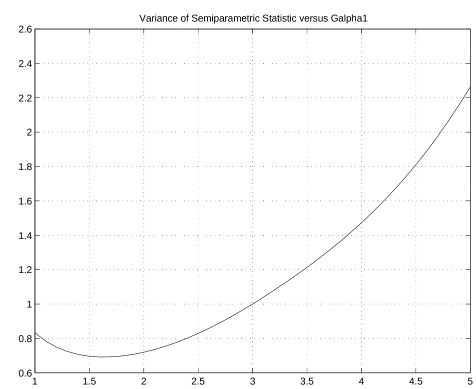

Figure 3: Variance of $\tilde{Z}^*$ versus $\alpha_{\gamma 1}$ when $X_1 \sim \text{Gamma}(\alpha_{\gamma 1}, 1)$, $X_2 \sim \text{Gamma}(3, 1)$.

2.2.1.3 Log Normal Example In this section, another example of the asymptotic $\tilde{\mathrm{Z}}^*$ distribution is calculated where X_1 and X_2 have log normal distributions with different μ_{l1} and μ_{l2} parameters and with a common $\sigma_l^2 = 1$ parameter as described in section 2.1.1.4.

Concerning the convergence conditions of Assumption 2.1, $h(x)$ is continuous and non-constant with respect to the log normal density, and $h^k(x) = \log^k(x)$ integrable with respect to the log normal density for $k = 1, 2, 3, 4$, given

$$\mathrm{E}\left(h^k\left(X_j\right)\right) = \mathrm{E}\left(Y_j^k\right) \text{ where } X_j \sim \mathrm{LN}\left(\mu_{lj}, \sigma_l^2\right),\ Y_j \sim \mathrm{N}\left(\mu_{lj}, \sigma_l^2\right)$$

and using the moments identified in section 2.1.1.1. Hence the conditions are met that allow $(\hat{\alpha}, \hat{\beta})$ to converge in probability to their true value (α_0, β_0). Also the convergence results of (13), (14), (15), (16), (17) are valid. In conclusion, $\tilde{\mathrm{Z}}_n^*$ converges in distribution to $\tilde{\mathrm{Z}}^*$ identified in (44). Figure 4 graphs the variance of $\tilde{\mathrm{Z}}^*$ versus a range of μ_{l1} parameter values for X_1 with $\mu_{l2} = 0$ for X_2.

Concerning the convergence conditions of Assumption 2.2, the log normal density $p(x|\mu_l, \sigma_l^2)$ is a continuous function of its parameters (μ_l, σ_l^2) such that $g_1(x) = p(x|\mu_{l1}, \sigma_l^2) \to g_2(x) = p(x|\mu_{l2}, \sigma_l^2)$ for all $x \in \mathbb{R}^+$ as $\mu_{l1} \to \mu_{l2}$. The distortion parameters (α_n, β_n) are also continuous functions of the log normal parameters (μ_l, σ_l^2) as identified in section 2.1.1.4 such that $(\alpha_n, \beta_n) \to (0, 0)$ as $\mu_{l1} \to \mu_{l2}$. The function $h(x) = \log(x)$ is continuous and non-constant with respect to the log normal density, $h^k(x) = \log^k(x)$ is integrable with respect to the sequence of log normal densities for $k \in \{1, 2, 3, 4\}$ as identified above, and $E_n \log^4(X_1) \to E_0 \log^4(X_1)$ as $\mu_{l1} \to \mu_{l2}$. Hence the convergence conditions have been met that allow $\tilde{\mathrm{Z}}_n^*$ to converge in distribution to $\tilde{\mathrm{Z}}^* \sim \mathrm{N}(0, 1)$ as $\beta_n \to 0$.

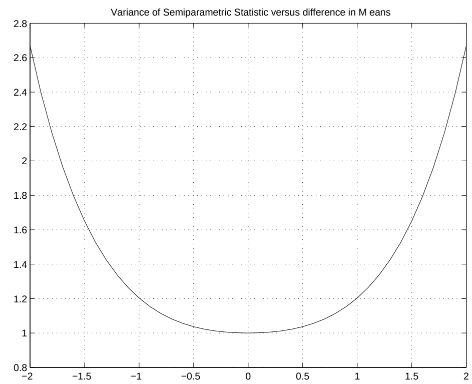

Figure 4: Variance of $\tilde{\mathrm{Z}}^*$ versus μ_{l1} when $X_1 \sim \mathrm{LN}(\mu_{l1}, 1)$, $X_2 \sim \mathrm{LN}(0, 1)$.

2.2.1.4 Limiting Example as (α_0, β_0) approaches 0 In this section, the limiting distribution for a sequence of $\tilde{\mathrm{Z}}^*$ random variables is calculated as (α_0, β_0) approaches $\mathbf{0}$.

$$\boldsymbol{D}_1 \to \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sigma_h \end{pmatrix}, \ \boldsymbol{\Sigma}_3 \to \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \ \mathbf{V}_1 \to \frac{1}{\sigma_h^2} \begin{pmatrix} \mu_h^2 & -\mu_h \\ -\mu_h & 1 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_1 \to \frac{\rho_1}{(1+\rho_1)^2} \begin{pmatrix} \sigma_h^2 & \mu_{h^3} - \mu_h \mu_{h^2} \\ \mu_{h^3} - \mu_h \mu_{h^2} & \mu_{h^4} - \mu_{h^2}^2 \end{pmatrix}$$

$$\boldsymbol{D}_1' \boldsymbol{M} \boldsymbol{\Sigma} \boldsymbol{M} \boldsymbol{D}_1 \to \begin{pmatrix} 0 & 0 \end{pmatrix} \boldsymbol{\Sigma}_1 \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & \sigma_h \end{pmatrix} \mathbf{V}_1 \begin{pmatrix} 0 \\ \sigma_h \end{pmatrix} = 1 \qquad (45)$$

The previous display shows the distribution of $\tilde{\mathrm{Z}}^*$ approaching a $\mathrm{N}(0, 1)$ distribution as (α_0, β_0) approaches $\mathbf{0}$. This result is expected since the original $\tilde{\mathrm{Z}}_n$ statistic converges to a $\mathrm{N}(0, 1)$ random variable when (α_0, β_0) equals $\mathbf{0}$, see section 2.1 (6).

2.2.2 Asymptotic Distribution of the T Statistic

In this section the asymptotic distribution is found for the common T statistic. In the first subsection, the independent random samples are assumed to be distributed according to two Gaussian densities with different means and with a common variance. In subsequent subsections, this Gaussian assumption is relaxed. Let T_n^2 rename the T^2 random variable defined by Cassela and Berger (1990) [6] in Theorem 11.2.2 for the case $k = 2$. Let $\boldsymbol{x}_1 = (x_{11}, \ldots, x_{1n_1})'$ represent a random sample from X_1. Let $\boldsymbol{x}_2 = (x_{21}, \ldots, x_{2n_2})'$ represent a random sample from X_2 independent of $\boldsymbol{x}_1$.

$$
\begin{aligned}
&x_{11}, \ldots, x_{1n_1} \sim X_1 \text{ with } g_1(x) = \left(\mu_1, \sigma_1^2\right) \text{ pdf} \\
&x_{21}, \ldots, x_{2n_2} \sim X_2 \text{ with } g_2(x) = \left(\mu_2, \sigma_2^2\right) \text{ pdf} \\
&\mathrm{T}_n^2 = \frac{n_1\left((\bar{x}_{1\cdot} - \bar{x}_{\cdot\cdot}) - (\mu_1 - \bar{\mu}_\cdot)\right)^2 + n_2\left((\bar{x}_{2\cdot} - \bar{x}_{\cdot\cdot}) - (\mu_2 - \bar{\mu}_\cdot)\right)^2}{\mathrm{S}_p^2} \\
&\mathrm{S}_p^2 = \frac{1}{n-2}\left((n_1 - 1)\mathrm{S}_1^2 + (n_2 - 1)\mathrm{S}_2^2\right) \\
&\mathrm{S}_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(x_{ji} - \bar{x}_{j\cdot})^2 \text{ for } j = 1, 2 \\
&\bar{x}_{j\cdot} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji} \text{ for } j = 1, 2 \text{ and } \bar{x}_{\cdot\cdot} = \frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j} x_{ji}
\end{aligned}
$$

For the case where X_1 and X_2 have Gaussian distributions with a common variance, then T_n^2 follows an F distribution with $(1, n-2)$ degrees of freedom, and T_n follows a student t distribution with $(n-2)$ degrees of freedom

$$
\begin{gathered}
X_1 \sim \mathrm{N}(\mu_1, \sigma^2) \text{ and } X_1 \sim \mathrm{N}(\mu_1, \sigma^2) \\
\mathrm{T}_n^2 \sim F_{1,n-2} \text{ and } \mathrm{T}_n \sim t_{n-2}\ .
\end{gathered}
$$

After a little algebra, the T_n random variable is rewritten as

$$
\mathrm{T}_n = \frac{\sqrt{\frac{1}{1+\rho_1}}\sqrt{n_1}\left(\bar{x}_{1\cdot} - \mu_1\right) - \sqrt{\frac{\rho_1}{1+\rho_1}}\sqrt{n_2}\left(\bar{x}_{2\cdot} - \mu_2\right)}{\mathrm{S}_p}\ .
$$

Under the null hypothesis $\mathbf{H}_0 : \mu_1 = \mu_2$, the T_n random variable becomes

$$
\mathrm{T}_{0n} \equiv \sqrt{\frac{n_1 n_2}{n}}\left(\frac{\bar{x}_{1\cdot} - \bar{x}_{2\cdot}}{\mathrm{S}_p}\right)\ .
$$

2.2.2.1 Asymptotics of T Statistic Assuming Normality In this section, the independent random samples are assumed to come from two Gaussian densities with different means $\mu_1 \neq \mu_2$ and with a common variance σ^2.

$$X_1 \sim \mathrm{N}\left(\mu_1, \sigma^2\right), \ X_2 \sim \mathrm{N}\left(\mu_2, \sigma^2\right)$$

Lemma 2.16. *If $\boldsymbol{x}_1$, a random sample from $X_1 \sim N(\mu_1, \sigma^2)$, is independent of $\boldsymbol{x}_2$, a random sample from $X_2 \sim N(\mu_2, \sigma^2)$, then T_n converges in distribution to a standard Gaussian random variable $N(0,1)$.*

Proof: The asymptotic distribution of T_n is found by using the independence property of X_1 and X_2, by applying the law of large numbers, by applying the continuous mapping theorem, and by applying Slutsky's theorem

$$\begin{pmatrix} \sqrt{n_1}\left(\bar{x}_{1\cdot} - \mu_1\right) \\ \sqrt{n_2}\left(\bar{x}_{2\cdot} - \mu_2\right) \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_2\right) \xrightarrow{d} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_2\right)$$

$$\mathrm{S}_p^2 \xrightarrow{P} \sigma^2 \text{ and } \mathrm{T}_n \xrightarrow{d} \mathrm{T}_* = \frac{\sqrt{\frac{1}{1+\rho_1}} Z_1 - \sqrt{\frac{\rho_1}{1+\rho_1}} Z_2}{\sigma} \sim \mathrm{N}\left(0,1\right). \blacksquare \qquad (46)$$

Under the null hypothesis $\mathbf{H}_0 : \mu_1 = \mu_2$, the T_{0n} statistic also converges to a standard Gaussian random variable

$$\mathrm{T}_{0n} \equiv \sqrt{\frac{n_1 n_2}{n}} \left(\frac{\bar{x}_{1\cdot} - \bar{x}_{2\cdot}}{\mathrm{S}_p}\right) \xrightarrow{d} \mathrm{T}_0 \sim \mathrm{N}(0,1) \ . \qquad (47)$$

As shown in sections 2.2.2.2 and 2.2.2.2.1, the multivariate central limit theorem is applied to find the asymptotic distribution of T_{0n} minus a suitable offset under the conditions of the alternative hypothesis $\mathbf{H}_1 : \mu_1 \neq \mu_2$ when X_1 and X_2 are not necessarily Gaussian. In this section a direct approach, that does not rely on the multivariate central limit theorem, is used to find the asymptotic distribution of T_{0n} minus a suitable offset when X_1 and X_2 are Gaussian. Also in this section, the mean and variance for the offset T_{0n} statistic is shown to converge to the mean and variance of the asymptotic distribution for the offset T_{0n} statistic under the Gaussian assumption.

In the following display, the T_{0n} statistic minus a suitable offset, is rewritten as the linear combination of three random variables

$$
\begin{aligned}
\text{Let } \mathrm{T}_{0n}^{*} &\equiv \sqrt{\frac{n_1 n_2}{n}} \left(\frac{\bar{x}_{1\cdot} - \bar{x}_{2\cdot}}{\mathrm{S}_p} - \frac{\mu_1 - \mu_2}{\sigma} \right) \\
&= \frac{1}{\mathrm{S}_p} \sqrt{\frac{1}{1+\rho_1}} \sqrt{n_1} \left(\bar{x}_{1\cdot} - \mu_1 \right) - \frac{1}{\mathrm{S}_p} \sqrt{\frac{\rho_1}{1+\rho_1}} \sqrt{n_2} \left(\bar{x}_{2\cdot} - \mu_2 \right) \\
&- \left(\frac{\mu_1 - \mu_2}{\mathrm{S}_p^2 \sigma + \mathrm{S}_p \sigma^2} \right) \sqrt{\frac{\rho_1}{(1+\rho_1)^2}} \sqrt{n} \left(\mathrm{S}_p^2 - \sigma^2 \right)
\end{aligned}
$$

which is rewritten in vector notation as $\mathrm{T}_{0n}^{*} = \boldsymbol{D}_n' \boldsymbol{Y}_n$

$$
\boldsymbol{D}_n \equiv \begin{pmatrix} \frac{1}{\mathrm{S}_p} \sqrt{\frac{1}{1+\rho_1}} \\ -\frac{1}{\mathrm{S}_p} \sqrt{\frac{\rho_1}{1+\rho_1}} \\ -\left(\frac{\mu_1 - \mu_2}{\mathrm{S}_p^2 \sigma + \mathrm{S}_p \sigma^2} \right) \sqrt{\frac{\rho_1}{(1+\rho_1)^2}} \end{pmatrix}, \; \boldsymbol{Y}_n = \begin{pmatrix} y_{1n} \\ y_{2n} \\ y_{3n} \end{pmatrix} \equiv \begin{pmatrix} \sqrt{n_1} \left(\bar{x}_{1\cdot} - \mu_1 \right) \\ \sqrt{n_2} \left(\bar{x}_{2\cdot} - \mu_2 \right) \\ \sqrt{n} \left(\mathrm{S}_p^2 - \sigma^2 \right) \end{pmatrix} .
$$

For convenience of notation, the components in the decomposition of T_{0n}^{*} are denoted as $(\boldsymbol{D}_n, \boldsymbol{Y}_n)$. The components of $(\boldsymbol{D}_n, \boldsymbol{Y}_n)$ represent stochastic quantities that are different from the identically labeled components in the decomposition of $\tilde{\mathrm{Z}}_n^{*}$, see (10) and (11). In other words, the symbols $\boldsymbol{D}_n$ and $\boldsymbol{Y}_n$ are overloaded.

Lemma 2.17. *If $\boldsymbol{x}_1$, a random sample from $X_1 \sim N(\mu_1, \sigma^2)$, is independent of $\boldsymbol{x}_2$, a random sample from $X_2 \sim N(\mu_2, \sigma^2)$, then T_{0n}^{*} converges in distribution to a Gaussian random variable T_0^{*}.*

Proof: Section 2.1.1.1 identifies the first four moments for X_j. The law of large numbers and the continuous mapping theorem are applied to find the asymptotic limit for S_p^2 and S_p, since X_j^k is integrable for $j = 1, 2$ and $k = 1, 2$. Hence, $\boldsymbol{D}_n$ converges to $\boldsymbol{D}$. Since y_{1n} and y_{2n} are independent, the joint distribution for $(y_{1n}, y_{2n})'$ is the product of the marginal distributions for y_{1n} and for y_{2n}. The bivariate Gaussian distribution for $(y_{1n}, y_{2n})'$ remains the same for all n, while the marginal distribution for y_{3n} evolves with n

$$\begin{aligned}
\mathrm{S}_p^2 &\to \sigma^2 \text{ and } \mathrm{S}_p \to \sigma \\
\boldsymbol{D}_n \to \boldsymbol{D} &= \begin{pmatrix} \frac{1}{\sigma}\sqrt{\frac{1}{1+\rho_1}} & -\frac{1}{\sigma}\sqrt{\frac{\rho_1}{1+\rho_1}} & -\frac{\mu_1-\mu_2}{2\sigma^3}\sqrt{\frac{\rho_1}{(1+\rho_1)^2}} \end{pmatrix}' \\
\begin{pmatrix} y_{1n} \\ y_{2n} \end{pmatrix} &\equiv \begin{pmatrix} \sqrt{n_1}\left(\bar{x}_{1\cdot}-\mu_1\right) \\ \sqrt{n_2}\left(\bar{x}_{2\cdot}-\mu_2\right) \end{pmatrix} \sim \mathrm{N}\left(0,\sigma^2\mathbf{I}_2\right) \xrightarrow{d} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathrm{N}\left(0,\sigma^2\mathbf{I}_2\right) \\
(n_j-1)\frac{S_j^2}{\sigma^2} &\sim \mathrm{Gamma}\left(\frac{n_j-1}{2},2\right),\ j=1,2 \\
\sqrt{n}S_p^2 &\sim \mathrm{Gamma}\left(\frac{n-2}{2},\frac{\sqrt{n}}{n-2}2\sigma^2\right) \\
y_{3n} &\equiv \sqrt{n}\left(\mathrm{S}_p^2-\sigma^2\right) \sim \left(0,\frac{n}{n-2}2\sigma^4\right) \to \left(0,2\sigma^4\right)\ .
\end{aligned}$$

The asymptotic distribution for y_{3n} is found by using the moment generating function for y_{3n}

$$\begin{aligned}
M_{y_{3n}}(t) &= \left(1-\frac{\sqrt{n}}{n-2}2\sigma^2 t\right)^{-\frac{n-2}{2}} e^{-\sqrt{n}\sigma^2 t},\ t < \left(\frac{n-2}{\sqrt{n}}\right)\frac{1}{2\sigma^2} \\
\log M_{y_{3n}}(t) &= -\sqrt{n}\sigma^2 t - \frac{n-2}{2}\log\left(1-\frac{\sqrt{n}}{n-2}2\sigma^2 t\right) \\
&= -\sqrt{n}\sigma^2 t + \frac{n-2}{2}\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{\sqrt{n}}{n-2}2\sigma^2 t\right)^k,\ \left|\frac{\sqrt{n}}{n-2}2\sigma^2 t\right| < 1 \\
&= \frac{n}{n-2}\sigma^4 t^2 + R_n(t) \\
R_n(t) &\equiv \frac{n-2}{2}\sum_{k=3}^{\infty}\frac{1}{k}\left(\frac{\sqrt{n}}{n-2}2\sigma^2 t\right)^k,\ \left|\frac{\sqrt{n}}{n-2}2\sigma^2 t\right| < 1 \\
&= \frac{n-2}{2}\left(\frac{\sqrt{n}}{n-2}2\sigma^2 t\right)^3\sum_{k=3}^{\infty}\frac{1}{k}\left(\frac{\sqrt{n}}{n-2}2\sigma^2 t\right)^{k-3}\ .
\end{aligned}$$

The remainder term $R_n(t)$ converges to zero, so that the mgf for y_{3n} converges to the mgf for a Gaussian random variable for all t in a neighborhood of zero. Hence, by the convergence of mgfs theorem 2.3.4 [6], y_{3n} converges in distribution to a Gaussian random variable

$$
\begin{aligned}
|R_n(t)| &\le \frac{n^{\frac{3}{2}}}{(n-2)^2} 4\sigma^6 t^3 \sum_{k=0}^{\infty} \left| \frac{\sqrt{n}}{n-2} 2\sigma^2 t \right|^k, \ |t| < \frac{n-2}{\sqrt{n}} \frac{1}{2\sigma^2} \\
&= \frac{n^{\frac{3}{2}}}{(n-2)^2} 4\sigma^6 t^3 \left(1 - \left| \frac{\sqrt{n}}{n-2} 2\sigma^2 t \right| \right)^{-1} \\
&\to \frac{0}{1} = 0 \\
M_{y_{3n}} &\to e^{\sigma^4 t^2}, \ t \in (-\infty, \infty) \\
y_{3n} &\xrightarrow{d} y_3 \sim \mathrm{N}\left(0, 2\sigma^4\right) \ .
\end{aligned}
$$

Each of the components of $\boldsymbol{Y}_n$ are independent, since $\bar{x}_1.$ and S_1^2 are independent and $\bar{x}_2.$ and S_2^2 are independent. So that the distribution of $\boldsymbol{Y}_n$ is the product of the joint distribution for $(y_{1n}, y_{2n})'$ and the marginal distribution for y_{3n}. Hence, the distribution of $\boldsymbol{Y}_n$ converges to a multivariate Gaussian distribution. The asymptotic distribution for T_{0n}^* follows by applying Slutsky's theorem

$$
\begin{aligned}
\boldsymbol{D}_n &\to \boldsymbol{D} = \left(\tfrac{1}{\sigma}\sqrt{\tfrac{1}{1+\rho_1}} \quad -\tfrac{1}{\sigma}\sqrt{\tfrac{\rho_1}{1+\rho_1}} \quad -\tfrac{\mu_1-\mu_2}{2\sigma^3}\sqrt{\tfrac{\rho_1}{(1+\rho_1)^2}} \right)' \\
\boldsymbol{Y}_n &\xrightarrow{d} \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2 & & \\ & \sigma^2 & \\ & & 2\sigma^4 \end{bmatrix} \right) \\
\mathrm{T}_{0n}^* &\xrightarrow{d} \mathrm{T}_0^* = \boldsymbol{D}'\boldsymbol{Y} \sim \mathrm{N}\left(0, 1 + \frac{1}{2}\frac{\rho_1}{(1+\rho_1)^2}\frac{(\mu_1-\mu_2)^2}{\sigma^2}\right). \blacksquare
\end{aligned}
\tag{48}
$$

For convenience of notation, the components in the decomposition of the asymptotic random variable T_0^* are denoted as $(\boldsymbol{D}, \boldsymbol{Y})$. The components of $(\boldsymbol{D}, \boldsymbol{Y})$ represent random variables that are different from the identically labeled components in the decomposition of the asymptotic random variable $\tilde{\mathrm{Z}}^*$, see (44).

As a check of (48), it is possible to show directly that $(\mathrm{E}(\mathrm{T}^*_{0n}), \mathrm{Var}(\mathrm{T}^*_{0n}))$ converges to $(0, 1 + \frac{1}{2}\frac{\rho_1}{(1+\rho_1)^2}\frac{(\mu_1-\mu_2)^2}{\sigma^2})$.

Proposition 2.2. *If $X_1 \sim N(\mu_1, \sigma^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$ are independent then $E(T^*_{0n})$ converges to zero.*

Proof: First, the mean of $(\bar{x}_1. - \bar{x}_2.)/\mathrm{S}_p$ is found, using the independence of the elements of $(\bar{x}_1., \bar{x}_2., \mathrm{S}_p^2)$

$$\mathrm{E}\left(\frac{\bar{x}_1. - \bar{x}_2.}{\mathrm{S}_p}\right) = \frac{\mu_1 - \mu_2}{\sigma}\left(\frac{n-2}{2}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \tag{49}$$

and then the limit is found as $n \to \infty$ using Stirling's gamma approximation.

Let $p \equiv \dfrac{n}{2} - 2$ and $p_* \equiv \dfrac{n-1}{2} - 2$

$$\begin{aligned}
\left(\frac{n-2}{2}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} &= (p+1)^{\frac{1}{2}}\frac{\Gamma(p_*+1)}{\Gamma(p+1)} \qquad (50)\\
&= \mathrm{e}^{\frac{1}{2}}(p+1)^{\frac{1}{2}}\frac{p_*^{p_*+\frac{1}{2}}}{p^{p+\frac{1}{2}}}\frac{\Gamma(p_*+1)}{\sqrt{2\pi}\mathrm{e}^{-p_*}p_*^{p_*+\frac{1}{2}}}\frac{\sqrt{2\pi}\mathrm{e}^{-p}p^{p+\frac{1}{2}}}{\Gamma(p+1)}\\
&= \mathrm{e}^{\frac{1}{2}}\left(\frac{n-2}{n-4}\right)^{\frac{1}{2}}\left(1 - \frac{\frac{1}{2}}{\frac{n}{2}-2}\right)^{\frac{n}{2}-2}\frac{\Gamma(p_*+1)}{\sqrt{2\pi}\mathrm{e}^{-p_*}p_*^{p_*+\frac{1}{2}}}\frac{\sqrt{2\pi}\mathrm{e}^{-p}p^{p+\frac{1}{2}}}{\Gamma(p+1)}\\
&\xrightarrow{n} \mathrm{e}^{\frac{1}{2}} \times 1 \times \mathrm{e}^{-\frac{1}{2}} \times 1 \times 1 = 1
\end{aligned}$$

and $\mathrm{E}\left(\dfrac{\bar{x}_1. - \bar{x}_2.}{\mathrm{S}_p}\right) \xrightarrow{n} \dfrac{\mu_1 - \mu_2}{\sigma}$ where $\xrightarrow{n}$ is shorthand for $\xrightarrow{n\uparrow\infty}$.

It will be shown in Lemma 2.18 that

$$\begin{aligned}
&n\left(\left(\frac{n-2}{2}\right)\frac{\Gamma^2\left(\frac{n-3}{2}\right)}{\Gamma^2\left(\frac{n-2}{2}\right)} - 1\right) \xrightarrow{n} \frac{3}{2}\\
&= n\left(\left(\frac{n-2}{2}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} - 1\right)\left(\left(\frac{n-2}{2}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} + 1\right) .
\end{aligned}$$

The previous display and equation (50) are used to show

$$\begin{aligned}
\left(\left(\frac{n-2}{2}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} + 1\right) &\xrightarrow{n} 2\\
n\left(\left(\frac{n-2}{2}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} - 1\right) &\xrightarrow{n} \frac{3}{4} .
\end{aligned}$$

Hence, the result is proven that

$$\mathrm{E}\left(\mathrm{T}_{0n}^{*}\right) = \frac{1}{\sqrt{n}}\left(\frac{\mu_1-\mu_2}{\sigma}\right)\left(\frac{\rho_1}{(1+\rho_1)^2}\right)^{\frac{1}{2}} n\left(\left(\frac{n-2}{2}\right)^{\frac{1}{2}}\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}-1\right)$$
$$\xrightarrow{n} 0. \blacksquare$$

Lemma 2.18.

$$n\left(\left(\frac{n-2}{2}\right)\frac{\Gamma^2\left(\frac{n-3}{2}\right)}{\Gamma^2\left(\frac{n-2}{2}\right)}-1\right) \xrightarrow{n} \frac{3}{2}\,. \tag{51}$$

Proof: A change of variable and Stirling's gamma function approximation are used to analyze (51)

$$\text{Let } p \equiv \frac{n}{2}-2,\ p_* \equiv \frac{n-1}{2}-2,\ \text{and } q \equiv \frac{1}{p}$$

$$n\left(\left(\frac{n-2}{2}\right)\frac{\Gamma^2\left(\frac{n-3}{2}\right)}{\Gamma^2\left(\frac{n-2}{2}\right)}-1\right) \tag{52}$$

$$\approx 2\,(p+2)\left((p+1)\,\frac{\sqrt{2\pi}\mathrm{e}^{-2p_*}p_*^{2\left(p_*+\frac{1}{2}\right)}}{\sqrt{2\pi}\mathrm{e}^{-2p}p^{2\left(p+\frac{1}{2}\right)}}-1\right) \tag{53}$$

$$= 2\,(1+2q)\left(\frac{\left(1-\frac{q}{2}\right)^{\frac{2}{q}}\mathrm{e}^1-1}{q}\right)+2\,(1+2q)\left(1-\frac{q}{2}\right)^{\frac{2}{q}}\mathrm{e}^1 \tag{54}$$

$$= n\left(\mathrm{e}^1\left(\frac{n-2}{n-4}\right)\left(\frac{n-5}{n-4}\right)^{n-4}-1\right)\,. \tag{55}$$

It will be shown in Lemma 2.19 that

$$\lim_{n\to\infty}\frac{n\left(\left(\frac{n-2}{2}\right)\frac{\Gamma^2\left(\frac{n-3}{2}\right)}{\Gamma^2\left(\frac{n-2}{2}\right)}-1\right)}{n\left(\mathrm{e}^1\left(\frac{n-2}{n-4}\right)\left(\frac{n-5}{n-4}\right)^{n-4}-1\right)}=1$$

so that equation (55) is a large number approximation for (52).

With $(1-q/2)^{2/q} = \exp((2/q)\ln(1-q/2))$, L'hospital's rule shows

$$\begin{aligned}
\lim_{q\downarrow 0} \frac{\left(1-\frac{q}{2}\right)^{\frac{2}{q}} \mathrm{e}^1 - 1}{q} &= \lim_{q\downarrow 0} -\mathrm{e}^1 \left(1-\frac{q}{2}\right)^{\frac{2}{q}} \frac{2\left(1-\frac{q}{2}\right)\ln\left(1-\frac{q}{2}\right)+q}{q^2-\frac{1}{2}q^3} \\
&= -\mathrm{e}^1\mathrm{e}^{-1} \lim_{q\downarrow 0} \frac{2\left(1-\frac{q}{2}\right)\ln\left(1-\frac{q}{2}\right)+q}{q^2-\frac{1}{2}q^3} \\
&= -\lim_{q\downarrow 0} \frac{-\ln\left(1-\frac{q}{2}\right)}{2q-\frac{3}{2}q^2} \\
&= -\lim_{q\downarrow 0} \frac{\frac{1}{2}\left(1-\frac{q}{2}\right)^{-1}}{2-3q} \\
&= -\frac{1}{4}\,.
\end{aligned}$$

Hence, the previous display converges to the desired result as $q \downarrow 0$

$$(54) \rightarrow 2\times 1\times\left(-\frac{1}{4}\right) + 2\times 1\times \mathrm{e}^{-1}\mathrm{e}^{1} = -\frac{1}{2}+2 = \frac{3}{2}. \blacksquare \qquad (56)$$

Use of Stirling's gamma function approximation in (53) is appropriate due to the following result.

Lemma 2.19. *Equation (55) is a large number approximation for (52).*

Proof: The following bounds on Stirling's gamma function approximation are taken from Rao (1973) [23] 1e.7

$$\begin{aligned}
\mathrm{e}^{\frac{1}{12\left(\frac{n-4}{2}\right)}} &< \frac{\Gamma\left(\frac{n-5}{2}+1\right)}{\sqrt{2\pi}\left(\frac{n-5}{2}\right)^{\frac{n-4}{2}}\mathrm{e}^{-\frac{n-5}{2}}} < \mathrm{e}^{\frac{1}{12\left(\frac{n-5}{2}\right)}} \\
\mathrm{e}^{\frac{1}{12\left(\frac{n-3}{2}\right)}} &< \frac{\Gamma\left(\frac{n-4}{2}+1\right)}{\sqrt{2\pi}\left(\frac{n-4}{2}\right)^{\frac{n-3}{2}}\mathrm{e}^{-\frac{n-4}{2}}} < \mathrm{e}^{\frac{1}{12\left(\frac{n-4}{2}\right)}}\,.
\end{aligned}$$

After some algebra, the previous display leads to

$$\begin{aligned}
1 &< \frac{n\left(\left(\frac{n-2}{2}\right)\frac{\Gamma^2\left(\frac{n-3}{2}\right)}{\Gamma^2\left(\frac{n-2}{2}\right)}-1\right)}{n\left(\mathrm{e}^1\left(\frac{n-2}{n-4}\right)\left(\frac{n-5}{n-4}\right)^{n-4}-1\right)} < 1+\frac{\mathrm{R}_n}{\mathrm{S}_n} \\
\mathrm{R}_n &\equiv n\left(\mathrm{e}^{\frac{1}{3(n-5)}-\frac{1}{3(n-3)}}-1\right) \\
\mathrm{S}_n &\equiv n\left(\mathrm{e}^1\left(\frac{n-2}{n-4}\right)\left(\frac{n-5}{n-4}\right)^{n-4}-1\right)\,.
\end{aligned}$$

The limit for R_n is found by using L'hospital's rule. The limit for S_n was previously found, see (55) and (56).

$$\text{Let } q \equiv \frac{1}{n}$$

$$\lim_{n\to\infty} \mathrm{R}_n = \lim_{q\downarrow 0} \frac{\mathrm{e}^{\frac{1}{3}\left(\frac{1}{5}\left(\frac{1}{1-5q}\right)-\frac{1}{5}-\frac{1}{3}\left(\frac{1}{1-3q}\right)+\frac{1}{3}\right)}-1}{q}$$

$$= \lim_{q\downarrow 0} \mathrm{e}^{\frac{1}{3}\left(\frac{1}{5}\left(\frac{1}{1-5q}\right)-\frac{1}{5}-\frac{1}{3}\left(\frac{1}{1-3q}\right)+\frac{1}{3}\right)} \frac{1}{3}\left(\left(\frac{1}{1-5q}\right)^2-\left(\frac{1}{1-3q}\right)^2\right)$$

$$= 1 \times \frac{1}{3} \times (1-1) = 0$$

$$\lim_{n\to\infty} \mathrm{S}_n = \frac{3}{2}$$

Hence (53) is a large number approximation for (52) since

$$\lim_{n\to\infty} \frac{n\left(\left(\frac{n-2}{2}\right)\frac{\Gamma^2\left(\frac{n-3}{2}\right)}{\Gamma^2\left(\frac{n-2}{2}\right)}-1\right)}{n\left(\mathrm{e}^1\left(\frac{n-2}{n-4}\right)\left(\frac{n-5}{n-4}\right)^{n-4}-1\right)} = 1. \blacksquare$$

Proposition 2.3. *If $X_1 \sim N(\mu_1, \sigma^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$ are independent then the variance of T^*_{0n} converges to the variance of T^*_0.*

Proof: The result is proven using the previous results of (49) from Proposition 2.2 and of (51) from Lemma 2.18.

$$\mathrm{E}\left(\frac{\bar{x}_{1\cdot}-\bar{x}_{2\cdot}}{\mathrm{S}_p}\right)^2 = \mathrm{E}\left(\bar{x}_{1\cdot}-\bar{x}_{2\cdot}\right)^2 \mathrm{E}\left(\frac{1}{\mathrm{S}_p^2}\right)$$

$$= \frac{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\sigma^2+\left(\mu_1-\mu_2\right)^2}{\sigma^2}\left(\frac{n-2}{n-4}\right) \to \frac{\left(\mu_1-\mu_2\right)^2}{\sigma^2}$$

$$\mathrm{Var}\left(\mathrm{T}^*_{0n}\right) = \left(\frac{n_1 n_2}{n}\right)\mathrm{Var}\left(\frac{\bar{x}_{1\cdot}-\bar{x}_{2\cdot}}{\mathrm{S}_p}\right)$$

$$= \left(1+\frac{2}{n-4}\right)+2\frac{\rho_1}{\left(1+\rho_1\right)^2}\frac{\left(\mu_1-\mu_2\right)^2}{\sigma^2}\frac{n}{n-4}$$

$$+ \frac{\rho_1}{\left(1+\rho_1\right)^2}\frac{\left(\mu_1-\mu_2\right)^2}{\sigma^2}n\left(1-\left(\frac{n-2}{2}\right)\frac{\Gamma^2\left(\frac{n-3}{2}\right)}{\Gamma^2\left(\frac{n-2}{2}\right)}\right)$$

$$\to 1+\frac{1}{2}\frac{\rho_1}{\left(1+\rho_1\right)^2}\frac{\left(\mu_1-\mu_2\right)^2}{\sigma^2}. \blacksquare$$

2.2.2.1.1 Gaussian Example

In this section, an example of the asymptotic T_0^* variance is calculated where X_1 and X_2 have Gaussian distributions with different means μ_1 and μ_2, and with a common variance $\sigma^2 = 1$ as described in section 2.1.1.1. Figure 5 graphs the variance of T_0^* versus the difference in means of X_1 and X_2.

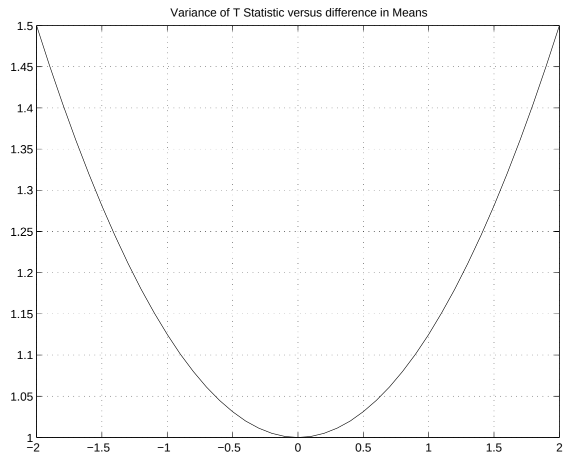

Figure 5: Variance of T_0^* versus $\mu_1 - \mu_2$ when $X_1 \sim N(\mu_1, 1)$, $X_2 \sim N(\mu_2, 1)$.

2.2.2.2 Asymptotics of T Statistic Without Assuming Normality

In this section, the independent random samples are assumed to come from two distributions, not necessarily Gaussian, with finite mean and variance

$$x_{11}, \ldots, x_{1n_1} \sim X_1 \text{ with } g_1(x) = \left(\mu_1, \sigma_1^2\right) \text{ pdf}$$
$$x_{21}, \ldots, x_{2n_2} \sim X_2 \text{ with } g_2(x) = \left(\mu_2, \sigma_2^2\right) \text{ pdf} \quad .$$

With these assumptions, the asymptotic distribution of T_{0n}, minus a suitable constant, is found under the conditions of the alternative hypothesis $\mathbf{H}_1 : \mu_1 \neq \mu_2$.

$$\text{Let } T_{0n}^* \equiv \sqrt{\frac{n_1 n_2}{n}} \left(\frac{\bar{x}_1. - \bar{x}_2.}{S_p} - \frac{\mu_1 - \mu_2}{\sigma_p} \right)$$
$$\sigma_p^2 \equiv \frac{\rho_1}{1+\rho_1}\sigma_1^2 + \frac{1}{1+\rho_1}\sigma_2^2$$

By direct linear expansion, T^*_{0n} is expressed as the linear combination of four random variables and a bias term

$$\begin{aligned}
\mathrm{T}^*_{0n} &= \sqrt{\frac{n_1 n_2}{n}} \left(\frac{\bar{x}_{1\cdot} - \mu_1}{\mathrm{S}_p} - \frac{\bar{x}_{2\cdot} - \mu_2}{\mathrm{S}_p} + \frac{\mu_1 - \mu_2}{\sigma_p \mathrm{S}_p} \left(\sigma_p - \mathrm{S}_p\right) \right) \\
&= \sqrt{\frac{n_1 n_2}{n}} \left(\frac{\bar{x}_{1\cdot} - \mu_1}{\mathrm{S}_p} - \frac{\bar{x}_{2\cdot} - \mu_2}{\mathrm{S}_p} \right) \\
&+ \sqrt{\frac{n_1 n_2}{n}} \left(\frac{n_1}{n}\sigma_1^2 - \frac{n_1 - 1}{n-2}\mathrm{S}_1^2 + \frac{n_2}{n}\sigma_2^2 - \frac{n_2 - 1}{n-2}\mathrm{S}_2^2 \right) \frac{(\mu_1 - \mu_2)}{\sigma_p \mathrm{S}_p \left(\sigma_p + \mathrm{S}_p\right)} \\
&= \sqrt{n_1}\left(\bar{x}_{1\cdot} - \mu_1\right) D_{1n} + \sqrt{n_1}\left(\bar{x}^2_{1\cdot} - \mathrm{E}\left(X_1^2\right)\right) D_{2n} \\
&+ \sqrt{n_2}\left(\bar{x}_{2\cdot} - \mu_2\right) D_{3n} + \sqrt{n_2}\left(\bar{x}^2_{2\cdot} - \mathrm{E}\left(X_2^2\right)\right) D_{4n} + B_n
\end{aligned}$$

where the coefficients are defined as

$$\begin{aligned}
D_{1n} &\equiv \sqrt{\frac{1}{1+\rho_1}}\frac{1}{\mathrm{S}_p} + \sqrt{\frac{\rho_1^2}{(1+\rho_1)^3}} \frac{(\bar{x}_{1\cdot} + \mu_1)(\mu_1 - \mu_2)}{\sigma_p \mathrm{S}_p \left(\sigma_p + \mathrm{S}_p\right)} \left(\frac{n}{n-2}\right) \\
D_{2n} &\equiv -\sqrt{\frac{\rho_1^2}{(1+\rho_1)^3}} \frac{(\mu_1 - \mu_2)}{\sigma_p \mathrm{S}_p \left(\sigma_p + \mathrm{S}_p\right)} \left(\frac{n}{n-2}\right) \\
D_{3n} &\equiv -\sqrt{\frac{\rho_1}{1+\rho_1}}\frac{1}{\mathrm{S}_p} + \sqrt{\frac{\rho_1}{(1+\rho_1)^3}} \frac{(\bar{x}_{2\cdot} + \mu_2)(\mu_1 - \mu_2)}{\sigma_p \mathrm{S}_p \left(\sigma_p + \mathrm{S}_p\right)} \left(\frac{n}{n-2}\right) \\
D_{4n} &\equiv -\sqrt{\frac{\rho_1}{(1+\rho_1)^3}} \frac{(\mu_1 - \mu_2)}{\sigma_p \mathrm{S}_p \left(\sigma_p + \mathrm{S}_p\right)} \left(\frac{n}{n-2}\right)
\end{aligned}$$

and where the bias term is defined as

$$B_n \equiv -2\sqrt{\frac{\rho_1}{(1+\rho_1)^2}} \frac{\sigma_p (\mu_1 - \mu_2)}{\mathrm{S}_p \left(\sigma_p + \mathrm{S}_p\right)} \frac{\sqrt{n}}{(n-2)} .$$

T^*_{0n} is then written in vector notation.

$$\begin{aligned}
&\mathrm{T}^*_{0n} = \boldsymbol{D}'_n \boldsymbol{Y}_n + B_n \\
&\boldsymbol{D}_n \equiv \begin{pmatrix} D_{1n} \\ D_{2n} \\ D_{3n} \\ D_{4n} \end{pmatrix}, \boldsymbol{Y}_n \equiv \begin{pmatrix} y_{1n} \\ y_{2n} \\ y_{3n} \\ y_{4n} \end{pmatrix} = \begin{pmatrix} \sqrt{n_1}\left(\bar{x}_{1\cdot} - \mu_1\right) \\ \sqrt{n_1}\left(\bar{x}^2_{1\cdot} - \mathrm{E}\left(X_1^2\right)\right) \\ \sqrt{n_2}\left(\bar{x}_{2\cdot} - \mu_2\right) \\ \sqrt{n_2}\left(\bar{x}^2_{2\cdot} - \mathrm{E}\left(X_2^2\right)\right) \end{pmatrix}
\end{aligned} \tag{57}$$

It follows immediately that $\boldsymbol{Y}_n$ has a mean of $\mathbf{0}$. Assuming that the first four moments are finite for X_1 and X_2, then $\boldsymbol{Y}_n$ has a constant variance matrix for all n

$$\boldsymbol{Y}_n \sim (\mathbf{0}, \boldsymbol{\Sigma})\,,\ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix},\ \boldsymbol{\Sigma}_j = \mathbf{Var}\begin{pmatrix} X_j - \mathrm{E}\left(X_j\right) \\ X_j^2 - \mathrm{E}\left(X_j^2\right) \end{pmatrix} \tag{58}$$

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} \sigma_j^2 & \mathrm{E}\left(X_j^3\right) - \mu_j \mathrm{E}\left(X_j^2\right) \\ \mathrm{E}\left(X_j^3\right) - \mu_j \mathrm{E}\left(X_j^2\right) & \mathrm{E}\left(X_j^4\right) - \mathrm{E}^2\left(X_j^2\right) \end{bmatrix},\ j = 1, 2\ .$$

Lemma 2.20. *If the first two moments of X_1 and X_2 are finite, then $\boldsymbol{D}_n$ converges in probability to $\boldsymbol{D}$ and B_n converges in probability to zero.*

Proof: The law of large numbers and the continuous mapping theorem are applied to show

$$\frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji}^k \xrightarrow{P} \mathrm{E}X_j^k \text{ for } j \in \{1,2\},\ k \in \{1,2\}$$
$$\mathrm{S}_p \xrightarrow{P} \sigma_p\ .$$

The previous display is used to find the convergence in probability limit for the four coefficients in $\boldsymbol{D}_n$ and the bias term B_n/D_n^*, assuming the first two moments of X_1 and X_2 are finite

$$\boldsymbol{D}_n \xrightarrow{P} \boldsymbol{D} = \begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1}{1+\rho_1}}\frac{1}{\sigma_p} + \sqrt{\frac{\rho_1^2}{(1+\rho_1)^3}}\mu_1\frac{(\mu_1-\mu_2)}{\sigma_p^3} \\ -\sqrt{\frac{\rho_1^2}{(1+\rho_1)^3}}\frac{(\mu_1-\mu_2)}{2\sigma_p^3} \\ -\sqrt{\frac{\rho_1}{1+\rho_1}}\frac{1}{\sigma_p} + \sqrt{\frac{\rho_1}{(1+\rho_1)^3}}\mu_2\frac{(\mu_1-\mu_2)}{\sigma_p^3} \\ -\sqrt{\frac{\rho_1}{(1+\rho_1)^3}}\frac{(\mu_1-\mu_2)}{2\sigma_p^3} \end{pmatrix}$$
$$B_n \xrightarrow{P} -\sqrt{\frac{\rho_1}{(1+\rho_1)^2}}\frac{(\mu_1-\mu_2)}{\sigma_p} \times 0 = 0.\ \blacksquare$$

Lemma 2.21. *If the first four moments of X_1 and X_2 are finite, then the random vector $\boldsymbol{Y}_n$ converges in distribution to a multivariate Gaussian random vector $\boldsymbol{Y}$.*

Proof: The multivariate central limit theorem ([23], 2c.5) is applied to find the asymptotic distribution for the random vector $\boldsymbol{Y}_n$, assuming the first four moments of X_1 and X_2 are finite

$$\boldsymbol{Y}_n \xrightarrow{d} \boldsymbol{Y} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})\,,\ \mathbf{Var}(\boldsymbol{Y}_n) = \boldsymbol{\Sigma} = \mathbf{Var}(\boldsymbol{Y})$$

by showing every linear combination of $\boldsymbol{Y}_n$ converges in distribution to a univariate Gaussian distribution

$$z_n = \boldsymbol{\lambda}'\boldsymbol{Y}_n \xrightarrow{d} z = \boldsymbol{\lambda}'\boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}\right) \tag{59}$$
$$\boldsymbol{\lambda} = \left(\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2'\right)', \ \boldsymbol{\lambda}_1 = (\lambda_{11}, \lambda_{12},)', \ \boldsymbol{\lambda}_2 = (\lambda_{21}, \lambda_{22})' \ .$$

The Lindeberg-Feller form of the central limit theorem ([23], 2c.5) is applied to show (59).

$$\text{Let } z_{ji} = \frac{1}{\sqrt{\rho_j}}\boldsymbol{\lambda}_j' \begin{pmatrix} x_{ji} - \mathrm{E}\left(X_j\right) \\ x_{ji}^2 - \mathrm{E}\left(X_j^2\right) \end{pmatrix} \sim G_{z_{ji}} = G_{Z_j}, \ j = 1, 2, \ i = 1 \dots n_j$$
$$Z_j \sim \left(\mathrm{E}\left(Z_j\right), \mathrm{Var}\left(Z_j\right)\right) = \left(0, \frac{1}{\rho_j}\boldsymbol{\lambda}_j'\boldsymbol{\Sigma}_j\boldsymbol{\lambda}_j\right), \ j = 1, 2$$
$$\begin{aligned}\text{Let } C_n^2 &= \sum_{i=1}^{n_1} \mathrm{Var}\left(z_{1i}\right) + \sum_{i=1}^{n_2} \mathrm{Var}\left(z_{2i}\right) \\ &= \frac{n_1}{\rho_1}\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}_1\boldsymbol{\lambda}_1 + n_2\boldsymbol{\lambda}_2'\boldsymbol{\Sigma}_2\boldsymbol{\lambda}_2 \\ &= n_2\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}\end{aligned}$$

The Lindeberg-Feller convergence condition, as specialized to (59), is satisfied for any $\varepsilon > 0$

$$\begin{aligned}&\frac{1}{C_n^2}\left(\sum_{i=1}^{n_1}\int_{|z|>\varepsilon C_n} z^2 dG_{z_{1i}}(z) + \sum_{i=1}^{n_2}\int_{|z|>\varepsilon C_n} z^2 dG_{z_{2i}}(z)\right) \\ =&\frac{\rho_1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}\int I\left(|z| > \varepsilon C_n\right) z^2 dG_{Z_1}(z) + \frac{1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}\int I\left(|z| > \varepsilon C_n\right) z^2 dG_{Z_2}(z) \\ &\to 0 \text{ as } n \uparrow \infty\end{aligned}$$

since $\mathrm{Var}(z_n) = \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}$ is constant and finite for all n and since the convergence of the two integrals to zero follows by applying the dominated convergence theorem. Hence

$$\frac{\sum_{i=1}^{n_1} z_{1i} + \sum_{i=1}^{n_2} z_{2i}}{\sqrt{n_2\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}}} \xrightarrow{d} \mathrm{N}\left(0, 1\right)$$

which proves the result that

$$\begin{aligned}\boldsymbol{\lambda}'\boldsymbol{Y}_n &= \frac{\sqrt{\rho_1}}{\sqrt{n_1}}\sum_{i=1}^{n_1} z_{1i} + \frac{1}{\sqrt{n_2}}\sum_{i=1}^{n_2} z_{2i} = \frac{1}{\sqrt{n_2}}\left(\sum_{i=1}^{n_1} z_{1i} + \sum_{i=1}^{n_2} z_{2i}\right) \\ &\xrightarrow{d} \mathrm{N}\left(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}\right). \ \blacksquare\end{aligned}$$

In conclusion, Lemmas 2.20 and 2.21 are combined to find the asymptotic distribution for T_{0n}^*.

Theorem 2.5. *If the first four moments of X_1 and X_2 are finite, then T_{0n}^* converges in distribution to a Gaussian random variable T_0^*.*

Proof: The asymptotic distribution for T_{0n}^* is found, by applying the results of Lemmas 2.20 and 2.21, and by applying Slutsky's theorem

$$\mathrm{T}_{0n}^* = \boldsymbol{D}_n' \boldsymbol{Y}_n + B_n/D_n^* \xrightarrow{d} \mathrm{T}_0^* = \boldsymbol{D}'\boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D}\right). \blacksquare$$

In order to derive the Pitman efficiencies, the following results show that T_{0n}^* converges to a standard Gaussian distribution $\mathrm{T}_0^* \sim \mathrm{N}(0,1)$ if $g_1(x) = p_n(x) \to g_2(x)$ almost everywhere as $n \to \infty$. In the sequel, let the operators $\mathrm{E}_n(\cdot)$ and $\mathrm{Var}_n(\cdot)$ denote expectation and variance with respect to a density that varies with n.

Lemma 2.22. *Let $\{p_n(x) : n = 0, 1, 2, \dots\}$ define a sequence of density functions where $X_1 \sim p_n$ at time index n such that $p_n(x) \to p_0(x)$ almost everywhere. Let $X_2 \sim g_2$. If $E_n|X_1^k| \to E_0|X_1^k|$ for $k \in \{1,2\}$ and X_2^k is integrable for $k \in \{1,2\}$, then $\boldsymbol{D}_n$ converges in probability to $\boldsymbol{D}$ and B_n converges in probability to zero.*

Proof: At time index n, let $\{x_{ni} : i = 1, \dots, n_1\}$ denote a random sample from the probability distribution P_n associated with the density p_n. Proposition 2.1 with $f(x) = x$ and $f(x) = x^2$ and with $\rho = \rho_1$ shows that

$$\frac{1}{n_1}\sum_{i=1}^{n_1} x_{ni}^k \xrightarrow{P_n} \mathrm{E}_0 X_1^k, \text{ for } k \in \{1,2\}.$$

Let $x_{2i} \sim g_2$ for $i = 1, \dots, n_2$. The independent identically distributed version of the weak law of large numbers is applied to show

$$\frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i}^k \xrightarrow{P} \mathrm{E} X_2^k, \text{ for } k \in \{1,2\}, .$$

The two previous displays together with the continuous mapping theorem are used to show $\mathrm{S}_p \xrightarrow{P_n} \sigma_p$. The previous statement in combination with the two previous displays proves the result. $\blacksquare$

Lemma 2.23. *Let $\{p_n(x) : n = 0, 1, 2, \dots\}$ define a sequence of density functions where $X_1 \sim p_n$ at time index n such that $p_n(x) \to p_0(x)$ almost* *everywhere. Let $X_2 \sim g_2$. If $E_n|X_1^k| \to E_0|X_1^k|$ for $k \in \{1,2,3,4\}$ and X_2^k is integrable for $k \in \{1,2,3,4\}$, then the random vector $\boldsymbol{Y}_n$ converges in distribution to a multivariate Gaussian random vector $\boldsymbol{Y}$*

$$\boldsymbol{Y}_n \overset{d(P_n)}{\to} \boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$$
$$\mathbf{Var}_n(\boldsymbol{Y}_n) = \boldsymbol{\Sigma}_n \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{1n} & \\ & \boldsymbol{\Sigma}_2 \end{bmatrix} \to \boldsymbol{\Sigma}_0 \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{10} & \\ & \boldsymbol{\Sigma}_2 \end{bmatrix} = \mathbf{Var}_0(\boldsymbol{Y})$$

where $\boldsymbol{Y}_n$ remains as defined in (58) with EX_1^k replaced by $E_nX_1^k$ for $k \in \{1,2\}$, where $\boldsymbol{\Sigma}_{1n}$ and $\boldsymbol{\Sigma}_{10}$ have the same structure as $\boldsymbol{\Sigma}_1$ defined in (58) with EX_1^k replaced by $E_nX_1^k$ in $\boldsymbol{\Sigma}_{1n}$ for $k \in \{1,2,3,4\}$ and with EX_1^k replaced by $E_0X_1^k$ in $\boldsymbol{\Sigma}_{10}$ for $k \in \{1,2,3,4\}$, and where $\boldsymbol{\Sigma}_2$ remains the same as defined in (58).

Proof: As shown in Lemma 2.21, the multivariate central limit theorem ([23], 2c.5) is applied to show the convergence in joint distribution of $\boldsymbol{Y}_n$

$$z_n = \boldsymbol{\lambda}'\boldsymbol{Y}_n \overset{d(P_n)}{\to} z = \boldsymbol{\lambda}'\boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}_0\boldsymbol{\lambda}\right)$$
$$\boldsymbol{\lambda} = \left(\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2'\right)', \ \boldsymbol{\lambda}_1 = (\lambda_{11}, \lambda_{12},)', \ \boldsymbol{\lambda}_2 = (\lambda_{21}, \lambda_{22})' \ .$$

The Lindeberg-Feller form of the central limit theorem ([30] , Proposition 2.27) is applied to show the previous display. Let z_{ji} and C_n remain defined as in Lemma 2.21 such that for $i = 1, \ldots, n_j$ and $j = 1, 2$

$$\begin{aligned}
&z_{1i} \sim G_{n,z_{1i}} = G_{n,Z_1}, \ z_{2i} \sim G_{z_{2i}} = G_{Z_2} \\
&Z_1 \sim (\mathrm{E}_n(Z_1), \mathrm{Var}_n(Z_1)) = \left(0, \frac{1}{\rho_1}\boldsymbol{\lambda}_1'\boldsymbol{\Sigma}_{1n}\boldsymbol{\lambda}_1\right) \\
&Z_2 \sim (\mathrm{E}(Z_2), \mathrm{Var}(Z_2)) = \left(0, \frac{1}{\rho_2}\boldsymbol{\lambda}_2'\boldsymbol{\Sigma}_2\boldsymbol{\lambda}_2\right) \\
&C_n^2 \equiv \sum_{i=1}^{n_1} \mathrm{Var}_n(z_{1i}) + \sum_{i=1}^{n_2} \mathrm{Var}(z_{2i}) = n_2\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda} \ .
\end{aligned}$$

The Lindeberg-Feller convergence condition, as specialized to z_{ji}/C_n is satisfied for any $\varepsilon > 0$

$$\begin{aligned}
&\left(\sum_{i=1}^{n_1} \int_{\left|\frac{z}{C_n}\right|>\varepsilon} \left(\frac{z}{C_n}\right)^2 dG_{n,z_{1i}}(z) + \sum_{i=1}^{n_2} \int_{\left|\frac{z}{C_n}\right|>\varepsilon} \left(\frac{z}{C_n}\right)^2 dG_{z_{2i}}(z)\right) \\
=&\frac{\rho_1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}} \int I(|z| > \varepsilon C_n) z^2 dG_{n,Z_1}(z) + \frac{1}{\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}} \int I(|z| > \varepsilon C_n) z^2 dG_{Z_2}(z) \\
&\to 0 \text{ as } n \uparrow \infty
\end{aligned}$$

where

$$\sum_{i=1}^{n_1} \mathrm{Var}\frac{z_{1i}}{C_n} + \sum_{i=1}^{n_2} \mathrm{Var}\frac{z_{2i}}{C_n} = 1$$

since both integrals converge to zero by applying Pratt's extended dominated convergence theorem from Appendix 2B [23] with $\mathrm{Var}_n Z_1 \to \mathrm{Var}_0 Z_1 < \infty$ and with Var $Z_2 < \infty$ and since $\boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda} \to \boldsymbol{\lambda}'\boldsymbol{\Sigma}_0\boldsymbol{\lambda} < \infty$, hence

$$\frac{\sum_{i=1}^{n_1} z_{1i} + \sum_{i=1}^{n_2} z_{2i}}{\sqrt{n_2 \boldsymbol{\lambda}'\boldsymbol{\Sigma}_n\boldsymbol{\lambda}}} \overset{d(P_n)}{\to} \mathrm{N}\left(0, 1\right)$$

which proves the result that

$$\boldsymbol{\lambda}'\boldsymbol{Y}_n \overset{d(P_n)}{\to} \mathrm{N}\left(0, \boldsymbol{\lambda}'\boldsymbol{\Sigma}_0\boldsymbol{\lambda}\right) . \blacksquare$$

Theorem 2.6. *Let $\{p_n(x) : n = 0, 1, 2, \dots\}$ define a sequence of density functions where $X_1 \sim p_n$ at time index n such that $p_n(x) \to p_0(x)$ almost everywhere. Let $X_2 \sim g_2$. If $E_n|X_1^k| \to E_0|X_1^k|$ for $k \in \{1, 2, 3, 4\}$ and X_2^k is integrable for $k \in \{1, 2, 3, 4\}$, then T_{0n}^* converges in distribution to a Gaussian random variable T_0^*.*

Proof: The asymptotic distribution for T_{0n}^* is found, by applying the results of Lemmas 2.22 and 2.23, and by applying Slutsky's theorem

$$\mathrm{T}_{0n}^* = \boldsymbol{D}_n'\boldsymbol{Y}_n + B_n \overset{d(P_n)}{\to} \mathrm{T}_0^* = \boldsymbol{D}'\boldsymbol{Y} \sim \mathrm{N}\left(0, \boldsymbol{D}'\boldsymbol{\Sigma}_0\boldsymbol{D}\right) . \blacksquare$$

In order to satisfy the convergence conditions that $\mathrm{E}_n|X_1^k| \to \mathrm{E}_0|X_1^k|$ for $k \in \{1, 2, 3, 4\}$, it suffices to show that $\mathrm{E}_n X_1^4 \to \mathrm{E}_0 X_1^4$. The remaining moment convergence conditions are satisfied by applying Pratt's extended dominated convergence theorem from Appendix 2B [23] since $|x^k|$ for $k \in \{1, 2, 3\}$ is bounded by $1 + x^4$. In order to satisfy the integrable moment conditions on X_2^k for $k \in \{1, 2, 3, 4\}$, it suffices to show that X_2^4 is integrable. The remaining integrable moment conditions are satisfied since $\mathrm{E}X_2^4 < \infty$ implies that $\mathrm{E}|X_2^k| < \infty$ for $k \in \{1, 2, 3\}$ by applying the Lyapunov inequality.

Corollary 2.13. *If the limiting density $p_0(x)$ is the same as the reference density $g_2(x)$ then the limiting distribution of T_0^* is a standard Gaussian distribution: $T_0^* \sim N(0, 1)$.*

Proof: Under the assumptions where $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2 = \sigma_p^2$, direct calculation shows that

$$\boldsymbol{D} = \left(\sqrt{\tfrac{1}{1+\rho_1}}\tfrac{1}{\sigma_p}, \quad 0, \quad -\sqrt{\tfrac{\rho_1}{1+\rho_1}}\tfrac{1}{\sigma_p}, \quad 0\right)', \; \boldsymbol{\Sigma}_{10} = \boldsymbol{\Sigma}_2 = \begin{bmatrix} \sigma_p^2 & * \\ * & * \end{bmatrix}$$

Hence the result is proven since $\boldsymbol{D}'\boldsymbol{\Sigma}_0\boldsymbol{D} = 1$. ■

For convenience of notation, the components in the decomposition of the random variable T^*_{0n} are denoted as $(\boldsymbol{D}_n, \boldsymbol{Y}_n)$, and the components in the decomposition of the asymptotic random variable T^*_0 are denoted as $(\boldsymbol{D}, \boldsymbol{Y})$. The components of $(\boldsymbol{D}_n, \boldsymbol{Y}_n)$ and of $(\boldsymbol{D}, \boldsymbol{Y})$ are different from the identically labeled components in the decompositions of the random variable $\tilde{\mathrm{Z}}^*_n$ and of the asymptotic random variable $\tilde{\mathrm{Z}}^*$, see (44). In a similar manner, the covariance structure $\boldsymbol{\Sigma}$ of the random variable $\boldsymbol{Y}$ from the decomposition of the asymptotic random variable T^*_0 is different from the identically labeled covariance structure of the random variable $\boldsymbol{Y}$ from the decomposition of the asymptotic random variable $\tilde{\mathrm{Z}}^*$.

The next subsection shows that the variance of T^*_0 reduces to (48) under the Gaussian assumption.

2.2.2.2.1 Gaussian Example

In this section, an example of the asymptotic T_0^* distribution is examined where X_1 and X_2 have differing Gaussian distributions. The integrable moment condtions of Theorem 2.5 are satisfied since the Gaussian distribution has finite moments of all orders

$$\begin{aligned} X_j &\sim \mathrm{N}\left(\mu_j, \sigma_j^2\right), \text{ for } j = 1, 2 \\ \mathrm{E}\left(X_j^3\right) &= 2\sigma_j^2\mu_j \\ \mathrm{E}\left(X_j^4\right) &= 2\sigma_j^2\left(\sigma_j^2 + 2\mu_j^2\right) \ . \end{aligned}$$

The additional convergence conditions of Theorem 2.6 are also satisfied since the Gaussian density is a continuous function of its parameters such that the X_1 density $g_1(x) = \mathrm{N}(\mu_1, \sigma_1^2)$ converges to the X_2 density $g_2(x) = \mathrm{N}(\mu_2, \sigma_2^2)$ as $(\mu_1, \sigma_1^2) \rightarrow (\mu_2, \sigma_2^2)$ for all $x \in \mathbb{R}$ and since the fourth moment is a continuous function of the Gaussian parameters. The resulting variance for $\boldsymbol{Y}$ and distribution for T_0^* follow

$$\begin{aligned} \mathrm{Var}\left(\boldsymbol{Y}\right) &= \boldsymbol{\Sigma} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \\ \boldsymbol{\Sigma}_j &= \sigma_j^2 \begin{bmatrix} 1 & 2\mu_j \\ 2\mu_j & 2\left(\sigma_j^2 + 2\mu_j\right) \end{bmatrix}, \text{ for } j = 1, 2 \\ \mathrm{T}_0^* &= \boldsymbol{D}'\boldsymbol{Y} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D}\right) \\ \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D} &= \begin{pmatrix} D_1 & D_2 \end{pmatrix} \boldsymbol{\Sigma}_1 \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} + \begin{pmatrix} D_3 & D_4 \end{pmatrix} \boldsymbol{\Sigma}_2 \begin{pmatrix} D_3 \\ D_4 \end{pmatrix} \\ &= \frac{1}{1+\rho_1}\frac{\sigma_1^2}{\sigma_p^2}\left(1 + \frac{1}{2}\frac{\rho_1^2}{\left(1+\rho_1\right)^2}\frac{\sigma_1^2}{\sigma_p^4}\left(\mu_1 - \mu_2\right)^2\right) \\ &+ \frac{\rho_1}{1+\rho_1}\frac{\sigma_2^2}{\sigma_p^2}\left(1 + \frac{1}{2}\frac{1}{\left(1+\rho_1\right)^2}\frac{\sigma_2^2}{\sigma_p^4}\left(\mu_1 - \mu_2\right)^2\right) \\ &= \frac{\left(\sigma_1^2 + \rho_1\sigma_2^2\right)}{\left(\rho_1\sigma_1^2 + \sigma_2^2\right)} + \frac{1}{2}\rho_1\frac{\left(\rho_1\sigma_1^4 + \sigma_2^4\right)}{\left(\rho_1\sigma_1^2 + \sigma_2^2\right)^3}\left(\mu_1 - \mu_2\right)^2 \ . \end{aligned}$$

If X_1 and X_2 have common variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$ as described in section 2.1.1.1, then the resulting variance for T_0^* is consistent with previous results from (48)

$$\boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D} = 1 + \frac{1}{2}\frac{\rho_1}{\left(1+\rho_1\right)^2}\frac{\left(\mu_1 - \mu_2\right)^2}{\sigma^2} \ .$$

2.2.2.2.2 Gamma Examples I and II

In this section, two examples of the asymptotic T_0^* distribution are examined where X_1 and X_2 have differing gamma distributions. The integrable moment condtions of Theorem 2.5 are satisfied since the gamma distribution has finite moments of all orders

$$\begin{aligned}
X_j &\sim \mathrm{Gamma}\left(\alpha_{\gamma j}, \beta_{\gamma j}\right), \text{ for } j = 1, 2 \\
\mathrm{E}\left(X_j\right) &= \alpha_{\gamma j}\beta_{\gamma j} \\
\mathrm{E}\left(X_j^2\right) &= \alpha_{\gamma j}(\alpha_{\gamma j}+1)\beta_{\gamma j}^2 \\
\mathrm{E}\left(X_j^3\right) &= \alpha_{\gamma j}(\alpha_{\gamma j}+1)(\alpha_{\gamma j}+2)\beta_{\gamma j}^3 \\
\mathrm{E}\left(X_j^4\right) &= \alpha_{\gamma j}(\alpha_{\gamma j}+1)(\alpha_{\gamma j}+2)(\alpha_{\gamma j}+3)\beta_{\gamma j}^4 \ .
\end{aligned}$$

The additional convergence conditions of Theorem 2.6 are also satisfied since the gamma density is a continuous function of its parameters such that the X_1 density $g_1(x) = \mathrm{Gamma}(\alpha_{\gamma 1}, \beta_{\gamma 1})$ converges to the X_2 density $g_2(x) = \mathrm{Gamma}(\alpha_{\gamma 2}, \beta_{\gamma 2})$ as $(\alpha_{\gamma 1}, \beta_{\gamma 1}) \to (\alpha_{\gamma 2}, \beta_{\gamma 2})$ for all $x \in \mathbb{R}^+$ and since the fourth moment is a continuous function of the gamma parameters. The resulting variance for $\boldsymbol{Y}$ and distribution for T_0^* follow:

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{Y}\right) &= \boldsymbol{\Sigma} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \\
\boldsymbol{\Sigma}_j &= \alpha_{\gamma j}\beta_{\gamma j}^2 \begin{bmatrix} 1 & 2(\alpha_{\gamma j}+1)\beta_{\gamma j} \\ 2(\alpha_{\gamma j}+1)\beta_{\gamma j} & 2(\alpha_{\gamma j}+1)(2\alpha_{\gamma j}+3)\beta_{\gamma j}^2 \end{bmatrix}, \ j = 1, 2 \\
\mathrm{T}_0^* &= \boldsymbol{D}'\boldsymbol{Y} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D}\right) \\
\boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D} &= \begin{pmatrix} D_1 & D_2 \end{pmatrix} \boldsymbol{\Sigma}_1 \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} + \begin{pmatrix} D_3 & D_4 \end{pmatrix} \boldsymbol{\Sigma}_2 \begin{pmatrix} D_3 \\ D_4 \end{pmatrix}
\end{aligned}$$

$$= \frac{1}{1+\rho_1}\frac{\sigma_1^2}{\sigma_p^2}\left(1 - 2\beta_{\gamma 1}\frac{(\mu_1-\mu_2)}{\sigma_p^2}\frac{\rho_1}{1+\rho_1} + \beta_{\gamma 1}^2\left(\alpha_1+3\right)\frac{\left(\mu_1-\mu_2\right)^2}{2\sigma_p^4}\frac{\rho_1^2}{\left(1+\rho_1\right)^2}\right)$$

$$+\frac{\rho_1}{1+\rho_1}\frac{\sigma_2^2}{\sigma_p^2}\left(1 + 2\beta_{\gamma 2}\frac{(\mu_1-\mu_2)}{\sigma_p^2}\frac{1}{1+\rho_1} + \beta_{\gamma 2}^2\left(\alpha_2+3\right)\frac{\left(\mu_1-\mu_2\right)^2}{2\sigma_p^4}\frac{1}{\left(1+\rho_1\right)^2}\right)$$

$$\begin{aligned}
&= \frac{\left(\sigma_1^2+\rho_1\sigma_2^2\right)}{\left(\rho_1\sigma_1^2+\sigma_2^2\right)} - 2\rho_1\left(\sigma_1^2\beta_{\gamma 1} - \sigma_2^2\beta_{\gamma 2}\right)\frac{\left(\mu_1-\mu_2\right)}{\left(\rho_1\sigma_1^2+\sigma_2^2\right)^2} \\
&\quad + \rho_1\left(\left(\alpha_{\gamma 1}+3\right)\rho_1\sigma_1^2\beta_{\gamma 1}^2 + \left(\alpha_{\gamma 2}+3\right)\sigma_2^2\beta_{\gamma 2}^2\right)\frac{\left(\mu_1-\mu_2\right)^2}{2\left(\rho_1\sigma_1^2+\sigma_2^2\right)^3} \ .
\end{aligned}$$

For the Gamma I example, where the gamma distributions for X_1 and X_2 have a common shape parameter $\alpha_{\gamma 1} = \alpha_{\gamma 2} = \alpha_\gamma$, as described in section 2.1.1.2, the resulting variance for T_0^* is

$$\boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D} = \frac{\left(\beta_{\gamma 1}^2 + \rho_1 \beta_{\gamma 2}^2\right)}{\left(\rho_1 \beta_{\gamma 1}^2 + \beta_{\gamma 2}^2\right)} - 2\rho_1 \left(\beta_{\gamma 1}^3 - \beta_{\gamma 2}^3\right) \frac{\left(\beta_{\gamma 1} - \beta_{\gamma 2}\right)}{\left(\rho_1 \beta_{\gamma 1}^2 + \beta_{\gamma 2}^2\right)^2} \tag{60}$$
$$+ \rho_1 \left(\alpha_\gamma + 3\right) \left(\rho_1 \beta_{\gamma 1}^4 + \beta_{\gamma 2}^4\right) \frac{\left(\beta_{\gamma 1} - \beta_{\gamma 2}\right)^2}{2\left(\rho_1 \beta_{\gamma 1}^2 + \beta_{\gamma 2}^2\right)^3} .$$

Figure 6 graphs the variance of T_0^* versus a range of $\beta_{\gamma 1}$ parameter values for X_1, with $\beta_{\gamma 2} = 3$ for X_2, and with $\alpha_\gamma = 1$ for both X_1 and X_2.

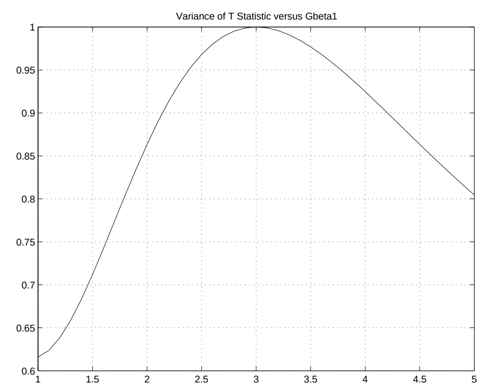

Figure 6: Variance of T_0^* versus $\beta_{\gamma 1}$ when $X_1 \sim \mathrm{Gamma}(1, \beta_{\gamma 1})$, $X_2 \sim \mathrm{Gamma}(1, 3)$.

For the Gamma II example, where the gamma distributions for X_1 and X_2 have a common scale parameter $\beta_{\gamma 1} = \beta_{\gamma 2} = \beta_\gamma$, as described in section 2.1.1.3, the resulting variance for T_0^* is

$$\boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D} = \frac{(\alpha_{\gamma 1} + \rho_1 \alpha_{\gamma 2})}{(\rho_1 \alpha_{\gamma 1} + \alpha_{\gamma 2})} + \rho_1 \left(\rho_1 \alpha_{\gamma 1} \left(\alpha_{\gamma 1} - 1\right) + \alpha_{\gamma 2} \left(\alpha_{\gamma 2} - 1\right)\right) \frac{\left(\alpha_{\gamma 1} - \alpha_{\gamma 2}\right)^2}{2 \left(\rho_1 \alpha_{\gamma 1} + \alpha_{\gamma 2}\right)^3} \,. \tag{61}$$

Figure 7 graphs the variance of T_0^* versus a range of $\alpha_{\gamma 1}$ parameter values for X_1, with $\alpha_{\gamma 2} = 3$ for X_2, and with $\beta_\gamma = 1$ for both X_1 and X_2.

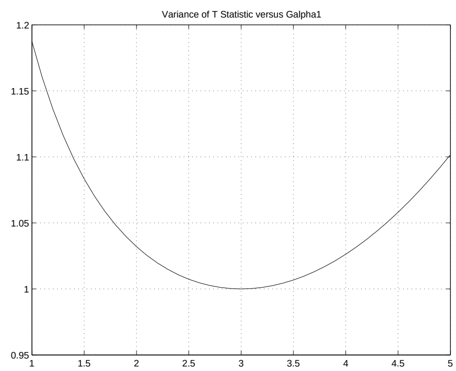

Figure 7: Variance of T_0^* versus $\alpha_{\gamma 1}$ when $X_1 \sim \text{Gamma}(\alpha_{\gamma 1}, 1)$, $X_2 \sim \text{Gamma}(3, 1)$.

2.2.2.2.3 Log Normal Example

In this section, another example of the asymptotic T_0^* distribution is examined where X_1 and X_2 have log normal distributions with different μ_{l1} and μ_{l2} parameters and with a common σ_l^2 parameter as described in section 2.1.1.4. The integrable moment condtions of Theorem 2.5 are satisfied since the log normal distribution has finite moments of all orders

$$X_j \sim \mathrm{LN}\left(\mu_{lj}, \sigma_l^2\right), \text{ for } j = 1, 2$$
$$\mathrm{E}\left(X_j^k\right) = e^{k\mu_{lj} + k^2\sigma_l^2/2}, \text{ for } k = 1, \ldots, 4 \, .$$

The additional convergence conditions of Theorem 2.6 are also satisfied since the log normal density is a continuous function of its parameters such that the X_1 density $g_1(x) = \mathrm{LN}(\mu_{l1}, \sigma_l^2)$ converges to the X_2 density $g_2(x) = \mathrm{LN}(\mu_{l2}, \sigma_l^2)$ as $\mu_{l1} \to \mu_{l2}$ for all $x \in \mathbb{R}^+$ and since the fourth moment is a continuous function of the log normal parameters. The resulting variance for $\boldsymbol{Y}$ and distribution for T_0^* follow:

$$\mathrm{Var}\left(\boldsymbol{Y}\right) = \boldsymbol{\Sigma} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}$$
$$\boldsymbol{\Sigma}_j = e^{2\mu_{lj} + \sigma_l^2} \begin{bmatrix} \left(e^{\sigma_l^2} - 1\right) & e^{\mu_{lj} + 3\sigma_l^2/2}\left(e^{2\sigma_l^2} - 1\right) \\ e^{\mu_{lj} + 3\sigma_l^2/2}\left(e^{2\sigma_l^2} - 1\right) & e^{2\mu_{lj} + 3\sigma_l^2}\left(e^{4\sigma_l^2} - 1\right) \end{bmatrix}, \ j = 1, 2$$
$$\mathrm{T}_0^* = \boldsymbol{D}'\boldsymbol{Y} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D}\right)$$
$$\boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D} = \begin{pmatrix} D_1 & D_2 \end{pmatrix} \boldsymbol{\Sigma}_1 \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} + \begin{pmatrix} D_3 & D_4 \end{pmatrix} \boldsymbol{\Sigma}_2 \begin{pmatrix} D_3 \\ D_4 \end{pmatrix} \, .$$

Figure 8 graphs the variance of T_0^* versus a range of μ_{l1} parameter values for X_1, with $\mu_{l2} = 0$ for X_2, and with $\sigma_l^2 = 1$ for both X_1 and X_2.

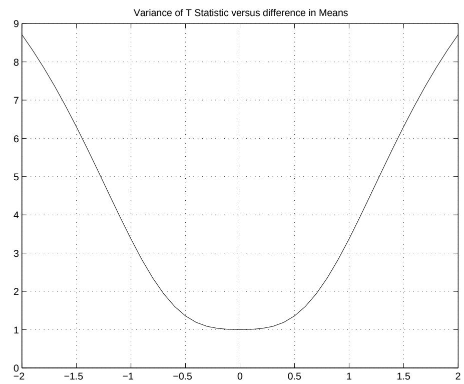

Figure 8: Variance of T_0^* versus μ_{l1} when $X_1 \sim \mathrm{LN}(\mu_{l1}, 1)$, $X_2 \sim \mathrm{LN}(0, 1)$.

2.2.2.2.4 Limiting Example as (μ_1, σ_1^2) Approaches (μ_2, σ_2^2)
In this section, the limiting distribution for a sequence of T_0^* random variables is found as (μ_1, σ_1^2) approaches (μ_2, σ_2^2). For this case, the variance of T_0^* approaches the limit in the following display.

$$\begin{aligned}\lim_{(\mu_1,\sigma_1^2)\to(\mu_2,\sigma_2^2)} \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D} &= \left(\frac{1}{1+\rho_1}\right)\left(\frac{1}{\sigma_2^2}\right)\left[\begin{pmatrix}1 & 0\end{pmatrix}\boldsymbol{\Sigma}_2\begin{pmatrix}1\\0\end{pmatrix}\right] \\ &+ \left(\frac{\rho_1}{1+\rho_1}\right)\left(\frac{1}{\sigma_2^2}\right)\left[\begin{pmatrix}1 & 0\end{pmatrix}\boldsymbol{\Sigma}_2\begin{pmatrix}1\\0\end{pmatrix}\right] \\ &= 1\end{aligned} \tag{62}$$

So that the distribution of T_0^* approaches a $\mathrm{N}(0,1)$ distribution as (μ_1, σ_1^2) approaches (μ_2, σ_2^2). This result is expected, since the original T_{0n} statistic converges to a $\mathrm{N}(0,1)$ random variable under the null hypothesis when $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ and when $(\mu_1, \sigma_1^2) = (\mu_2, \sigma_2^2)$, see (47). This result may be explicitly verified using the previous examples.

2.2.3 Relative Efficiency of T to $\tilde{\mathrm{Z}}_n$ Statistics

The $\tilde{\mathrm{Z}}_n$ statistic is used in testing the null hypothesis $\mathbf{H}_0 : \beta_0 = 0$. The T_{0n} statistic is used in testing the null hypothesis $\mathbf{H}_0 : \mu_1 = \mu_2$. Under the assumption that X_1 and X_2 are normally distributed with common variance

$$\beta_0 = (\mu_1 - \mu_2)/\sigma^2,$$

both of the statistics $\tilde{\mathrm{Z}}_n$ and T_{0n} are testing the null hypothesis that both of the normal distributions are the same.

This section uses relative efficiency and then Pitman efficiency, as described by Bickel and Doksum (1977) [2] 9.1.A, in order to compare the performance of the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests. Relative efficiency compares the sample sizes needed to achieve a desired power when the alternative hypothesis is true $\mathbf{H}_1 : \beta_0 \neq 0$ or equivalently $\mu_1 \neq \mu_2$. Since the $\tilde{\mathrm{Z}}^*_n$ random variable is asymptotically normal, a sample size $\mathrm{N}_z \gg 0$ is found to achieve a specified power for the $\tilde{\mathrm{Z}}_n$ test in terms of Φ. Also since the T^*_{0n} random variable is asymptotically normal, another sample size $\mathrm{N}_t \gg 0$ is also found to achieve a specified power for the T_{0n} test in terms of Φ. Let P_0 represent the probability distribution of the statistics when the null hypothesis is true. Let P_1 represent the probability distribution of the statistics when the alternative hypothesis is true. Then for the $\tilde{\mathrm{Z}}_n$ statistic

$$\begin{aligned}
&P_0\left(\left|\tilde{\mathrm{Z}}_{N_z}\right| > z(1-\alpha_{H0}/2)\right) \\
&\qquad \approx 1 - \Phi\left(z(1-\alpha_{H0}/2)\right) + \Phi\left(z(\alpha_{H0}/2)\right) = \alpha_{H0} \\
&P_1\left(\left|\tilde{\mathrm{Z}}_{N_z}\right| > z(1-\alpha_{H0}/2)\right) \\
&\qquad = P_1\left(\tilde{\mathrm{Z}}^*_{N_z} > z(1-\alpha_{H0}/2) - \sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sigma_h\beta_0\right) \\
&\qquad + P_1\left(\tilde{\mathrm{Z}}^*_{N_z} < -z(1-\alpha_{H0}/2) - \sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sigma_h\beta_0\right) \\
&\qquad \approx \Phi\left(\left(\sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sigma_h\beta_0 - z(1-\alpha_{H0}/2)\right)/\sigma\left(\tilde{\mathrm{Z}}^*\right)\right) && (63)\\
&\qquad + \Phi\left(\left(-\sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sigma_h\beta_0 - z(1-\alpha_{H0}/2)\right)/\sigma\left(\tilde{\mathrm{Z}}^*\right)\right) && (64)\\
&\quad \text{where } \sigma^2\left(\tilde{\mathrm{Z}}^*\right) = \mathrm{Var}\left(\tilde{\mathrm{Z}}^*\right)
\end{aligned}$$

and similarly for the T_{0n} statistic

$$
\begin{aligned}
&P_0\left(\left|\mathrm{T}_{0N_t}\right| > z(1-\alpha_{H0}/2)\right) \\
&\qquad \approx 1-\Phi\left(z(1-\alpha_{H0}/2)\right)+\Phi\left(z(\alpha_{H0}/2)\right)=\alpha_{H0} \\
&P_1\left(\left|\mathrm{T}_{0N_t}\right| > z(1-\alpha_{H0}/2)\right) \\
&\qquad = P_1\left(\mathrm{T}^*_{0N_t} > z(1-\alpha_{H0}/2)-\sqrt{\mathrm{N}_t}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\frac{(\mu_1-\mu_2)}{\sigma_p}\right) \\
&\qquad + P_1\left(\mathrm{T}^*_{0N_t} < -z(1-\alpha_{H0}/2)-\sqrt{\mathrm{N}_t}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\frac{(\mu_1-\mu_2)}{\sigma_p}\right) \\
&\qquad \approx \Phi\left(\left(\sqrt{\mathrm{N}_t}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\frac{(\mu_1-\mu_2)}{\sigma_p}-z(1-\alpha_{H0}/2)\right)/\sigma\left(\mathrm{T}^*_0\right)\right) \qquad (65) \\
&\qquad + \Phi\left(\left(-\sqrt{\mathrm{N}_t}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\frac{(\mu_1-\mu_2)}{\sigma_p}-z(1-\alpha_{H0}/2)\right)/\sigma\left(\mathrm{T}^*_0\right)\right) \qquad (66) \\
&\quad \text{where } \sigma^2\left(\mathrm{T}^*_0\right)=\mathrm{Var}\left(\mathrm{T}^*_0\right)\ .
\end{aligned}
$$

In order to compare the power of the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests, when the alternative hypothesis is true, it is natural to evaluate the ratio $\mathrm{N}_t/\mathrm{N}_z$, of the sample sizes needed to achieve a specific power value γ. As N_z and N_t grow, one of the power probabilities, from (63 or 64) and (65 or 66), increases to one; the other, to zero. Without loss of generality, assume that $\beta_0 > 0$ and $\mu_1 > \mu_2$ so that the first probability in each power, (63) and (65), increases to one as the sample size grows. Equating the power of the $\tilde{\mathrm{Z}}_n$ test to the power of the T_{0n} test, approximating the power of the $\tilde{\mathrm{Z}}_n$ test using (63), and approximating the power of the T_{0n} test using (65), leads to the following

$$
\begin{aligned}
P_1\left(\left|\tilde{\mathrm{Z}}_{N_z}\right| > z(1-\alpha_{H0}/2)\right) &= P_1\left(\left|\mathrm{T}_{0N_t}\right| > z(1-\alpha_{H0}/2)\right)=\gamma \\
\Phi^{-1}(\gamma)=z(\gamma) &= \frac{\sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sigma_h\beta_0-z(1-\alpha_{H0}/2)}{\sigma\left(\tilde{\mathrm{Z}}^*\right)} \qquad (67) \\
&= \frac{\sqrt{\mathrm{N}_t}\frac{\sqrt{\rho_1}}{(1+\rho_1)}\frac{(\mu_1-\mu_2)}{\sigma_p}-z(1-\alpha_{H0}/2)}{\sigma\left(\mathrm{T}^*_0\right)}\ . \qquad (68)
\end{aligned}
$$

For the case $0 < \gamma < 1$ where N_z and N_t are finite, equations (67) and (68) are used to calculate initial sample size approximations for N_z and N_t that give initial power values from (63)+(64) and (65)+(66) that are greater than or equal to the desired power value of γ due to (64) and (66). The correct sample sizes N_z and N_t are then found by decrementing the initial sample size approximations until (63) + (64) $\approx \gamma$ and (65) + (66) $\approx \gamma$. The ratio $\mathrm{N}_t/\mathrm{N}_z$, of the sample sizes needed to achieve a specific power, is called the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$.

For the case $\gamma \approx 1$ where $\mathrm{N}_z \gg 0$ and $\mathrm{N}_t \gg 0$, equating (67) and (68) leads to the following relative efficiency equations

$$\sqrt{\frac{\mathrm{N}_t}{\mathrm{N}_z}} = \sigma_h \beta_0 \frac{\sigma_p}{(\mu_1 - \mu_2)} \frac{\left(\sigma\left(\mathrm{T}_0^*\right) + z(1-\alpha_{H0}/2)/z\left(\gamma\right)\right)}{\left(\sigma\left(\tilde{\mathrm{Z}}^*\right) + z(1-\alpha_{H0}/2)/z\left(\gamma\right)\right)} \tag{69}$$

$$= \sigma_h \beta_0 \frac{\sigma_p}{(\mu_1 - \mu_2)} + \frac{z\left(\gamma\right)}{\sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}} \frac{\sigma_p}{(\mu_1 - \mu_2)} \left(\sigma\left(\mathrm{T}_0^*\right) - \sigma\left(\tilde{\mathrm{Z}}^*\right)\right) \tag{70}$$

$$= \sigma_h \beta_0 \frac{\sigma_p}{(\mu_1 - \mu_2)} \frac{\sigma\left(\mathrm{T}_0^*\right)}{\sigma\left(\tilde{\mathrm{Z}}^*\right)} + \frac{z(1-\alpha_{H0}/2)}{\sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}} \frac{\sigma_p}{(\mu_1 - \mu_2)} \left(1 - \frac{\sigma\left(\mathrm{T}_0^*\right)}{\sigma\left(\tilde{\mathrm{Z}}^*\right)}\right). \tag{71}$$

Either of the two relative efficiency equations, (69) or (71), is used to find the limit (if it exists) of the relative efficiency as γ increases to one while the other parameters are held constant. The limit of the relative efficiency as γ increases to one is called the asymptotic relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ or A.R.E. Van der Vaart (1998), in [30] section 8.2, provides an alternative limit definition for relative efficiency that is equivalent to the ratio $\sigma^2(\mathrm{T}_0^*)/\sigma^2(\tilde{\mathrm{Z}}^*)$.

Pitman efficiency, denoted as $e(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n)$, provides a way to compare the two test statistics, T_{0n} and $\tilde{\mathrm{Z}}_n$. Pitman efficiency is found by evaluating the ratio of sample sizes $\mathrm{N}_t/\mathrm{N}_z$ over a sequence of alternative hypotheses $\mathbf{H}_1 : \beta_0 \neq 0$ or equivalently $\mu_1 \neq \mu_2$ as $n \to \infty$ such that $\beta_0 \to 0$ and $\mu_1 \to \mu_2$ and such that the level value and power value of (67) and (68) remain fixed at α_{HO} and γ for each n. Requiring the power γ to remain constant implies that $\sqrt{\mathrm{N}_z}\beta_0 \to c_z \neq 0$ as $\mathrm{N}_z \to \infty$ and $\sqrt{\mathrm{N}_t}(\mu_1 - \mu_2)/\sigma_p \to c_t \neq 0$ as $\mathrm{N}_t \to \infty$, i.e. that the sequences $\beta_0 = O(1/\sqrt{n})$ and $(\mu_1 - \mu_2)/\sigma_p = O(1/\sqrt{n})$ as $n \to \infty$. In the Pitman efficiency analysis of the examples that follow, the moment functions $(\mu_1, \sigma_1^2) \equiv (\mu_1, \sigma_1^2)(\mathbf{\Theta}_{1n})$, $(\mu_2, \sigma_2^2) \equiv (\mu_2, \sigma_2^2)(\mathbf{\Theta}_2)$, and the distortion parameters $(\alpha_0, \beta_0) \equiv (\alpha_0, \beta_0)(\mathbf{\Theta}_{1n}, \mathbf{\Theta}_2) \equiv (\alpha_n, \beta_n)$, are functions of the distorted and reference densities parameters $g_1(x|\mathbf{\Theta}_{1n}) \equiv p_n(x)$ and $g_2(x|\mathbf{\Theta}_2)$, such that $\mathbf{\Theta}_2$ remains fixed and $g_1(x|\mathbf{\Theta}_{1n}) \to g_2(x|\mathbf{\Theta}_2)$ for every $x \in \mathbb{R}$ as $\mathbf{\Theta}_{1n} \to \mathbf{\Theta}_2$. Theorems 2.6 and 2.4 show that the convergence in distribution of T_{0n} to T_0^* and of $\tilde{\mathrm{Z}}_n$ to $\tilde{\mathrm{Z}}^*$ are valid as $n \to \infty$ when $p_n \to g_2$, $(\mu_1, \sigma_1^2)(\mathbf{\Theta}_{1n}) \to (\mu_2, \sigma_2^2)(\mathbf{\Theta}_2)$, $(\alpha_n, \beta_n) \to \mathbf{0}$ and under additional convergence conditions. In the examples that follow, only one of the distorted densities parameters $\theta_{1n} \in \mathbf{\Theta}_{1n}$ will vary such that $\mathbf{\Theta}_{1n} \to \mathbf{\Theta}_2$ as $\theta_{1n} \to \theta_2 \in \mathbf{\Theta}_2$. Let $\theta_{1n} \equiv \theta_2 + \theta_n$, $(\mu_1, \sigma_1^2)(\theta_n) \equiv (\mu_1, \sigma_1^2)(\mathbf{\Theta}_{1n})$, and

$(\alpha_0, \beta_0)(\theta_n) \equiv (\alpha_0, \beta_0)(\boldsymbol{\Theta}_{1n}, \boldsymbol{\Theta}_2)$. Note that the convergence in distribution properties of Theorems 2.6 and 2.4 are valid over any sequence $\theta_n \to 0$ as $n \to \infty$ which is a stronger result than just requiring the convergence in distribution properties to be valid over sequences $\theta_n = O(1/\sqrt{n})$. Theorem 14.19 from [30] provides a slope formula for the Pitman efficiency under conditions that are satisfied by the conditions and results of Theorems 2.6 and 2.4 with positive slopes $\mu'_T(0), \mu'_Z(0) > 0$ (see below). This theorem examines the Pitman efficiency over sequences $\theta_n = O(1/\sqrt{n})$. The Pitman efficiency slope formula is

$$\mu_T(\theta_n) \equiv \frac{\mu_1(\theta_n) - \mu_2}{\sigma_p(\theta_n)},\ \mu_Z(\theta_n) \equiv \sigma_h \beta_0(\theta_n)$$
$$e(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n) = \left(\frac{\mu'_Z(0)}{\mu'_T(0)}\right)^2 . \tag{72}$$

In equations (70) and (71), allowing $\sqrt{\mathrm{N}_z}(\mu_1(\theta_n) - \mu_2)$ to converge to a finite limit $c \neq 0$ as $n \to \infty$, such that both $\mu_1(\theta_n) - \mu_2$ and $\beta_0(\theta_n)$ converge to zero while $\sigma_p(\theta_n)$ converges to a finite positive constant and while α_{H0} and γ are held constant, results in the sample size ratio converging to a limit. It is easy to show that the Pitman efficiency slope formula (72) is equivalent to the limit from the previous statement as $\theta_n \to 0$ such that $\sqrt{\mathrm{N}_z}\theta_n \to c^* \neq 0$ with c^* finite and with α_{H0} and γ fixed

$$\frac{\mu'_Z(0)}{\mu'_T(0)} = \lim_{\theta_n \to 0} \frac{\mu_Z(\theta_n)/\theta_n}{\mu_T(\theta_n)/\theta_n} = \lim_{\theta_n \to 0} \frac{\mu_Z(\theta_n)}{\mu_T(\theta_n)} = \lim_{\sqrt{\mathrm{N}_z}(\mu_1(\theta_n)-\mu_2) \to c} \sqrt{\frac{\mathrm{N}_t}{\mathrm{N}_z}}$$

since as $\theta_n \to 0$ such that $\sqrt{\mathrm{N}_z}\theta_n \to c^* \neq 0$

$$\left(\mu_1, \sigma_1^2\right)(\theta_n) \to \left(\mu_2, \sigma_2^2\right),\ \sigma^2\left(\mathrm{T}_0^*\right) \to 1$$
$$\left(\alpha_0, \beta_0\right)(\theta_n) \to (0,0),\ \sigma^2(\tilde{\mathrm{Z}}^*) \to 1$$
$$\sqrt{\mathrm{N}_z}\left(\mu_1(\theta_n) - \mu_2\right) = \sqrt{\mathrm{N}_z}\theta_n \sigma_p(\theta_n)\frac{\mu_T(\theta_n)}{\theta_n} \to c^* \sigma_2 \mu'_T(0) = c\ .$$

Equation (67) is used to show as $\theta_n \to 0$

$$\sqrt{\mathrm{N}_z}\theta_n = \frac{\sigma\left(\tilde{\mathrm{Z}}^*\right) z(\gamma) + z(1 - \alpha_{H0}/2)}{\frac{\sqrt{\rho_1}}{(1+\rho_1)}\sigma_h \beta_0(\theta_n)}\theta_n = \frac{\sigma\left(\tilde{\mathrm{Z}}^*\right) z(\gamma) + z(1 - \alpha_{H0}/2)}{\frac{\sqrt{\rho_1}}{(1+\rho_1)}\mu_Z(\theta_n)/\theta_n}$$
$$\to \frac{z(\gamma) + z(1 - \alpha_{H0}/2)}{\frac{\sqrt{\rho_1}}{(1+\rho_1)}\mu'_Z(0)} = c^*\ .$$

2.2.3.1 Gaussian Example As an example, assume X_1 and X_2 have Gaussian distributions with different means $\mu_1 \neq \mu_2$ and with a common variance σ^2 as described in section 2.1.1.1. Note that $h(x) = x$. For this Gaussian example, the relative efficiency equation (71) is specialized to equation (73) below

$$\begin{aligned}
x_{11}, \ldots, x_{1n_1} &\sim X_1 \text{ with } \mathrm{N}\left(\mu_1, \sigma^2\right) \text{ pdf} \\
x_{21}, \ldots, x_{2n_1} &\sim X_2 \text{ with } \mathrm{N}\left(\mu_2, \sigma^2\right) \text{ pdf} \\
(\alpha_0, \beta_0) &= \left(\frac{\mu_2^2 - \mu_1^2}{2\sigma^2}, \frac{\mu_1 - \mu_2}{\sigma^2}\right), \; \sigma_h^2 = \sigma_p^2 = \sigma^2 \\
\frac{\sigma_p}{(\mu_1 - \mu_2)} &= \frac{\sigma}{(\mu_1 - \mu_2)}, \; \sigma_h \beta_0 \frac{\sigma_p}{(\mu_1 - \mu_2)} = 1 \\
\sqrt{\frac{\mathrm{N}_t}{\mathrm{N}_z}} &= \frac{\sigma\left(\mathrm{T}_0^*\right)}{\sigma\left(\tilde{\mathrm{Z}}^*\right)} + \frac{z(1 - \alpha_{H0})}{\sqrt{\mathrm{N}_z}\frac{\sqrt{\rho_1}}{(1+\rho_1)}} \frac{\sigma}{(\mu_1 - \mu_2)} \left(1 - \frac{\sigma\left(\mathrm{T}_0^*\right)}{\sigma\left(\tilde{\mathrm{Z}}^*\right)}\right)
\end{aligned} \tag{73}$$

Holding the distribution parameters (μ_1, μ_2, σ^2) constant, while increasing the power γ of the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests to one, results in the sample size ratio converging to the asymptotic relative efficiency of $\tilde{\mathrm{Z}}_n$ to T_{0n}.

$$\text{A.R.E.} \equiv \lim_{\gamma \to 1} \frac{\mathrm{N}_t}{\mathrm{N}_z} = \frac{\sigma^2\left(\mathrm{T}_0^*\right)}{\sigma^2\left(\tilde{\mathrm{Z}}^*\right)}$$

Allowing $\sqrt{\mathrm{N}_z}(\mu_1 - \mu_2)$ to converge to a finite limit $c \neq 0$, so that $\mu_1 - \mu_2$ and β_0 converge to zero with σ^2 constant, results in the sample size ratio converging to the Pitman efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$. The variance of T_0^* converges to one as μ_1 approaches μ_2 in (48). In general, as previously shown in (45), the variance of $\tilde{\mathrm{Z}}^*$ converges to one as β_0 approaches zero.

$$\begin{gathered}
\lim_{\beta_0 \to 0} \sigma^2\left(\tilde{\mathrm{Z}}^*\right) = 1, \; \lim_{(\mu_1 - \mu_2) \to 0} \sigma^2\left(\mathrm{T}_0^*\right) = 1 \\
e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) \equiv \lim_{\sqrt{\mathrm{N}_z}(\mu_1 - \mu_2) \to c} \frac{\mathrm{N}_t}{\mathrm{N}_z} = 1
\end{gathered}$$

Figures 1 and 5 graph the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* separately when $\sigma^2 = 1$. Figure 9 graphs the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* together when $\sigma^2 = 1$.

Figure 10 graphs the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ when $\sigma^2 = 1$, and when $\alpha_{H0} = .05$. In Figure 10, the relative efficiency is nearly one, in a neighborhood of $\mu_1 = \mu_2$. In other words, the sample sizes are approximately the same, for the T_{0n} and $\tilde{\mathrm{Z}}_n$ tests, in order to achieve the same power value, when the difference in means is small. Outside of this neighborhood of $\mu_1 = \mu_2$, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ decreases with larger power values. In other words, the T_{0n} test requires smaller random samples, relative to the semiparametric test, in order to achieve the same power value, as the power value increases.

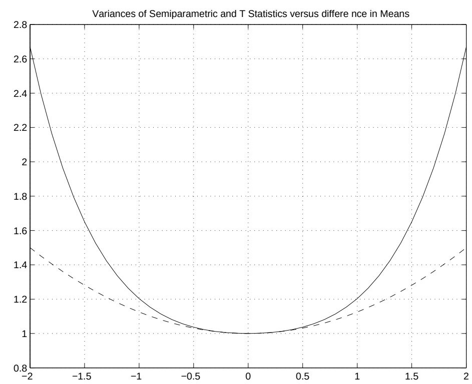

Figure 9: Given $X_1 \sim \mathrm{N}(\mu_1, 1)$, $X_2 \sim \mathrm{N}(\mu_2, 1)$, the solid line is the variance of $\tilde{\mathrm{Z}}^*$ versus $\mu_1 - \mu_2$, the dashed line is the variance of T_0^* versus $\mu_1 - \mu_2$.

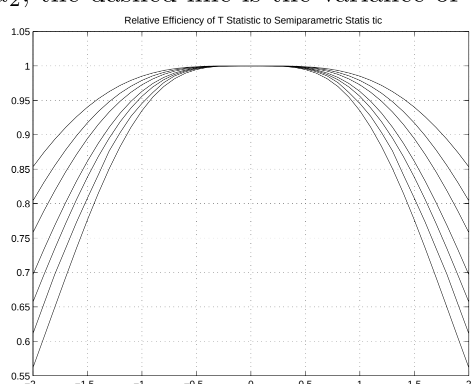

Figure 10: Relative Efficiency N_t/N_z curves of T_{0n} to $\tilde{\mathrm{Z}}_n$, versus $\mu_1 - \mu_2$, when $X_1 \sim \mathrm{N}(\mu_1, 1)$, $X_2 \sim \mathrm{N}(\mu_2, 1)$, and when $\alpha_{H0} = .05$. The curves, starting from the top, correspond to different power values of γ = .7, .8, .9, .99, .9999, .9999999999999999, 1.

The Pitman efficiency calculated using the slope formula (72) is consistent with the previous calculation. Let $\mu_{1n} \equiv \mu_2 + \theta_n \equiv \mu_1(\theta_n)$.

$$\mu_T(\theta_n) = \frac{\mu_1(\theta_n) - \mu_2}{\sigma_p(\theta_n)} = \frac{\theta_n}{\sigma},\ \mu_T'(0) = \frac{1}{\sigma}$$

$$\mu_Z(\theta_n) = \sigma_h \beta_0(\theta_n) = \frac{\theta_n}{\sigma},\ \mu_Z'(0) = \frac{1}{\sigma}$$

$$e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) = \left(\mu_Z'(0) / \mu_T'(0)\right)^2 = 1$$

Power simulation results for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests in Table 1 show how well the asymptotic power approximates finite sample behavior where $X_1 \sim \mathrm{N}(\mu_1, 1)$, $X_2 \sim \mathrm{N}(\mu_2, 1)$, and where $\mu_1 - \mu_2 = 0.2, 0.5$. Power simulation results for the T_{0n} and $\tilde{\mathrm{Z}}_n$ tests are also provided in Table 2 where $X_1 \sim \mathrm{N}(\mu_1, 1)$, $X_2 \sim \mathrm{N}(\mu_2, 1)$, and where $\mu_1 - \mu_2 = 1.0$. The combined Sample Sizes values $N_z = n_1 + n_2$ were calculated with $\rho_1 = 1$ via (67) and (63) + (64) to provide the specified $\alpha_{H0} = 0.05, 0.01$ error and to provide the specified Asymptotic Power values for $\tilde{\mathrm{Z}}_n$ that approximate a power of $\gamma = 0.80, 0.90$. The Asymptotic Power values for T_{0n} were calculated for the combined Sample Sizes values $N_t = n_1 + n_2$ with $\rho_1 = 1$ from (65) + (66). Relative Efficiency values were approximated using (69). A Relative Efficiency value less than one implies a larger Asymptotic Power value for the T_{0n} test versus the Asymptotic Power value for the $\tilde{\mathrm{Z}}_n$ test.

Table 1: Power simulation results for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests, using 500 independent runs, where $X_1 \sim \mathrm{N}(\mu_1, 1)$ and $X_2 \sim \mathrm{N}(\mu_2, 1)$, and where $\Delta\mu \equiv \mu_1 - \mu_2 = 0.2, 0.5$.

$\Delta\mu$	α_{H0}	Sample Sizes n_1, n_2	Sample Levels $\tilde{\mathrm{Z}}_n$, T_{0n}	Sample Powers $\tilde{\mathrm{Z}}_n$, T_{0n}	Asymptotic Powers $\tilde{\mathrm{Z}}_n(\gamma)$, T_{0n}	Rel. Eff. (69)
0.2	.05	394, 394	.046, .046	.800, .800	.8009, .8010	1.0000
0.2	.05	527, 527	.042, .042	.902, .902	.9003, .9003	0.9999
0.2	.01	585, 585	.010, .010	.808, .808	.8003, .8003	1.0000
0.2	.01	746, 746	.008, .008	.890, .888	.9003, .9004	1.0000
0.5	.05	64, 64	.048, .048	.824, .820	.8032, .8038	0.9984
0.5	.05	86, 86	.054, .052	.902, .900	.9024, .9030	0.9979
0.5	.01	95, 95	.010, .012	.804, .798	.8036, .8042	0.9987
0.5	.01	121, 121	.004, .004	.906, .904	.9014, .9020	0.9982

Table 2: Power simulation results for the $\tilde{Z}_n$ and T_{0n} tests, using 500 independent runs, where $X_1 \sim N(\mu_1, 1)$ and $X_2 \sim N(\mu_2, 1)$, and where $\Delta\mu \equiv \mu_1 - \mu_2 = 1.0$.

$\Delta\mu$	α_{H0}	Sample Sizes n_1, n_2	Sample Levels $\tilde{Z}_n$, T_{0n}	Sample Power $\tilde{Z}_n$, T_{0n}	Asymptotic Power $\tilde{Z}_n(\gamma)$, T_{0n}	Rel. Eff. (69)
1.0	.05	17, 17	.070, .070	.852, .850	.8081, .8162	0.9789
1.0	.05	23, 23	.058, .056	.934, .934	.9040, .9114	0.9726
1.0	.01	25, 25	.008, .006	.838, .826	.8092, .8172	0.9826
1.0	.01	32, 32	.012, .012	.908, .914	.9029, .9103	0.9768

In Tables 1 and 2, the important columns to compare are the Sample and Asymptotic Powers columns. The Sample Powers values for the $\tilde{Z}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level. The Sample Levels values for the $\tilde{Z}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level when the null hypothesis was true. For the simulations in Table 1, the Sample Sizes values are large enough so that the Sample Powers values are in agreement with the corresponding Asymptotic Powers values. For the simulations in Table 2, the Sample Sizes values are relatively small, so that some of the Sample Powers values are not quite in agreement with the corresponding Asymptotic Powers values. In both Tables 1 and 2, the Sample Powers values for the $\tilde{Z}_n$ and T_{0n} tests are nearly equal and are compatible with the Relative Efficiency values near 1.

The actual distribution of T_{0n}, when $X_1 \sim N(\mu_1, \sigma^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$, and when $\mu_1 \neq \mu_2$, is known to follow a non-central t distribution, with $n-2$ degrees of freedom, and with a non-centrality parameter δ

$$\delta = \sqrt{n}\frac{\sqrt{\rho_1}}{1+\rho_1}\left|\frac{\mu_1 - \mu_2}{\sigma}\right| .$$

For this example, it is interesting to compare the asymptotic power of T_{0n} against the true power of T_{0n}. Figures 11 and 12 graph the asymptotic power of T_{0n} versus $\phi = \delta/\sqrt{2}$ with $\alpha_{H0} = 0.05, 0.01$.

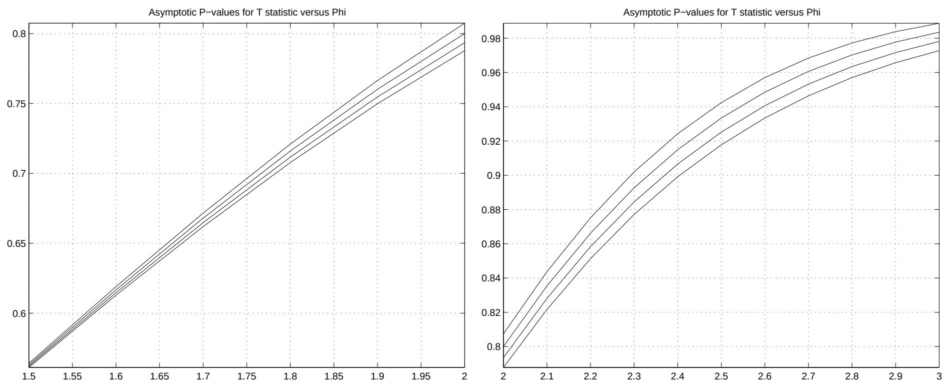

Figure 11: Asymptotic power of T_{0n} versus ϕ with $\alpha_{H0} = 0.05$. The curves, from the top, correspond to different degrees of freedom of $\nu = \infty, 60, 30, 20$.

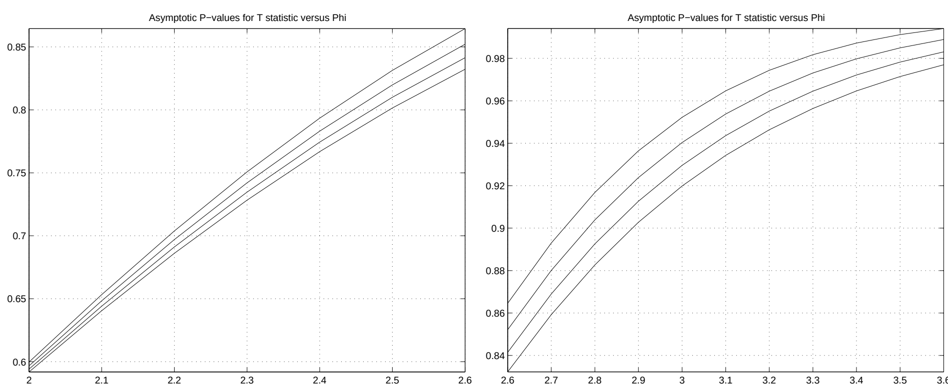

Figure 12: Asymptotic power of T_{0n} versus ϕ with $\alpha_{H0} = 0.01$. The curves, from the top, correspond to different degrees of freedom of $\nu = \infty, 60, 30, 20$.

Examination, of the $(\infty, 60)$ degrees of freedom curves in Figures 11 and 12, reveals that these curves are in close agreement with the corresponding curves in the Pearson and Hartley chart for the Power of the F tests found in Scheffe (1959) [27], where the numerator degrees of freedom is one. As expected, the other curves in Figures 11 and 12 with fewer degrees of freedom are in less agreement with the corresponding curves in the Pearson and Hartley chart, since the sample sizes are too small for the asymptotic distribution of T^*_{0n} to closely approximate the true distribution of T^*_{0n}.

2.2.3.2 Gamma Example I As another example, assume X_1 and X_2 have gamma distributions with a common shape parameter α_γ and with different scale parameters $\beta_{\gamma 1} \neq \beta_{\gamma 2}$ as described in section 2.1.1.2. Note that $h(x) = x$. For this Gamma Example I, the coefficients in the relative efficiency equation (71) are specialized to the coefficients in (74) and (75) below

$$\begin{aligned}
x_{11}, \ldots, x_{1n_1} &\sim X_1 \text{ with Gamma}\left(\alpha_\gamma, \beta_{\gamma 1}\right) \text{ pdf} \\
x_{21}, \ldots, x_{2n_2} &\sim X_2 \text{ with Gamma}\left(\alpha_\gamma, \beta_{\gamma 2}\right) \text{ pdf} \\
\left(\mu_j, \sigma_j^2\right) &= \left(\alpha_\gamma \beta_{\gamma j},\ \alpha_\gamma \beta_{\gamma j}^2\right),\ j = 1, 2 \\
(\alpha_0, \beta_0) &= \left(\alpha_\gamma \ln\left(\frac{\beta_{\gamma 2}}{\beta_{\gamma 1}}\right),\ \left(\frac{1}{\beta_{\gamma 2}} - \frac{1}{\beta_{\gamma 1}}\right)\right),\ \sigma_h^2 = \sigma_2^2 \\
\frac{\sigma_p}{(\mu_1 - \mu_2)} &= \frac{1}{(\beta_{\gamma 1} - \beta_{\gamma 2})} \sqrt{\frac{1}{\alpha_\gamma}\left(\frac{\rho_1 \beta_{\gamma 1}^2 + \beta_{\gamma 2}^2}{\rho_1 + 1}\right)} \qquad (74) \\
\sigma_h \beta_0 \frac{\sigma_p}{(\mu_1 - \mu_2)} &= \sqrt{\frac{(\beta_{\gamma 2}/\beta_{\gamma 1})^2 + \rho_1}{1 + \rho_1}}\,. \qquad (75)
\end{aligned}$$

The asymptotic relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ follows directly. With regard to the Pitman efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$, the variance of T_0^* converges to one as $\beta_{\gamma 1}$ approaches $\beta_{\gamma 2}$ in (60). In general, as previously shown in (45), the variance of $\tilde{\mathrm{Z}}^*$ converges to one as β_0 approaches zero.

$$\begin{aligned}
\text{A.R.E.} \equiv \lim_{\gamma \to 1} \frac{\mathrm{N}_t}{\mathrm{N}_z} &= \left(\frac{(\beta_{\gamma 2}/\beta_{\gamma 1})^2 + \rho_1}{1 + \rho_1}\right) \frac{\sigma^2\left(\mathrm{T}_0^*\right)}{\sigma^2\left(\tilde{\mathrm{Z}}^*\right)} \\
\lim_{\beta_0 \to 0} \sigma^2\left(\tilde{\mathrm{Z}}^*\right) = 1,\ &\lim_{(\beta_{\gamma 1} - \beta_{\gamma 2}) \to 0} \sigma^2\left(\mathrm{T}_0^*\right) = 1 \\
e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) &\equiv \lim_{\sqrt{\mathrm{N}_z}(\beta_{\gamma 1} - \beta_{\gamma 2}) \to c} \frac{\mathrm{N}_t}{\mathrm{N}_z} = 1
\end{aligned}$$

Figures 2 and 6 graph the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* separately when $\alpha_\gamma = 1$. Figure 13 graphs the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* together when $\alpha_\gamma = 1$. Figure 14 graphs the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ when $\alpha_\gamma = 1$, and when $\alpha_{H0} = .05$.

In Figure 14, the relative efficiency is nearly one, in a neighborhood of $\beta_{\gamma 1} = \beta_{\gamma 2} = 3$. Figure 14, also identifies some interesting relative efficiencies of T_{0n} to $\tilde{\mathrm{Z}}_n$, outside a neighborhood of $\beta_{\gamma 1} = \beta_{\gamma 2} = 3$. For smaller power values $\gamma = .7, .8, .9$, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ is greater than one, when $\beta_{\gamma 1} < \beta_{\gamma 2} = 3$. As the power value increases, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ decreases, so that at a power value of $\gamma = .99$, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ is less than one. In contrast, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ increases for $\beta_{\gamma 1} > \beta_{\gamma 2} = 3$ as the power values increase.

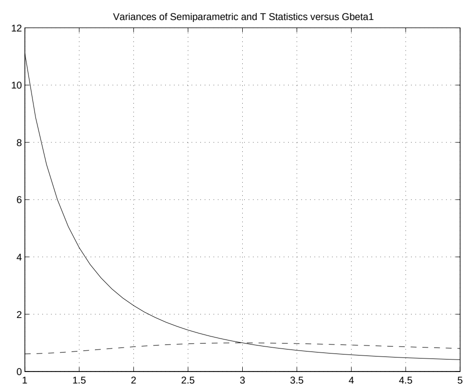

Figure 13: Given $X_1 \sim \mathrm{Gamma}(1, \beta_{\gamma 1})$, $X_2 \sim \mathrm{Gamma}(1, 3)$, solid line is the variance of $\tilde{\mathrm{Z}}^*$ versus $\beta_{\gamma 1}$, dashed line is the variance of T_0^* versus $\beta_{\gamma 1}$.

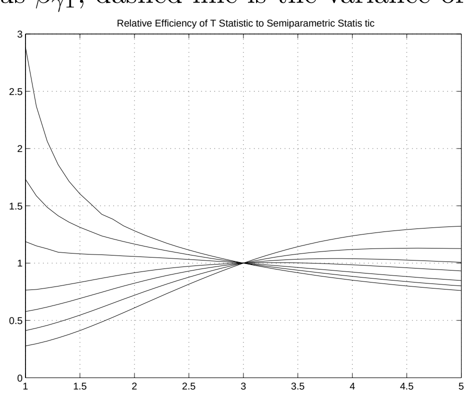

Figure 14: Relative Efficiency N_t/N_z curves of T_{0n} to $\tilde{\mathrm{Z}}_n$, versus $\beta_{\gamma 1}$, when $X_1 \sim \mathrm{Gamma}(1, \beta_{\gamma 1})$, $X_2 \sim \mathrm{Gamma}(1, 3)$, and when $\alpha_{H0} = .05$. The curves, starting from the top left, correspond to different power values of γ = .7, .8, .9, .99, .9999, .9999999999999999, 1.

The Pitman efficiency calculated using the slope formula (72) is consistent with the previous calculation. Let $\beta_{\gamma 1n} \equiv \beta_{\gamma 2} + \theta_n$.

$$\mu_T(\theta_n) = \frac{\mu_1(\theta_n) - \mu_2}{\sigma_p(\theta_n)} = \frac{\sqrt{\alpha_\gamma}\theta_n}{\left(\frac{\rho_1}{1+\rho_1}\beta_{\gamma 2}^2 + \frac{1}{1+\rho_1}(\beta_{\gamma 2} + \theta_n)^2\right)^{\frac{1}{2}}}$$

$$\mu_Z(\theta_n) = \sigma_h \beta_0(\theta_n) = \sqrt{\alpha_\gamma}\beta_{\gamma 2}\left(\frac{1}{\beta_{\gamma 2}} - \frac{1}{\beta_{\gamma 2} + \theta_n}\right)$$

$$\mu_T'(0) = \mu_Z'(0) = \frac{\sqrt{\alpha_\gamma}}{\beta_{\gamma 2}}$$

$$e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) = \left(\frac{\mu_Z'(0)}{\mu_T'(0)}\right)^2 = 1$$

Power simulation results for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests in Table 3 show how well the asymptotic power approximates finite sample behavior where $X_1 \sim \text{Gamma}(1, \beta_{\gamma 1})$, $X_2 \sim \text{Gamma}(1, 3)$, and where $\beta_{\gamma 1} = 2, 2.5, 3.5, 4$. The combined Sample Sizes values $N_z = n_1 + n_2$ were calculated with $\rho_1 = 1$ via (67) and (63) + (64) to provide the specified $\alpha_{H0} = 0.05, 0.01$ error and to provide the specified Asymptotic Power values for $\tilde{\mathrm{Z}}_n$ that approximate a power of $\gamma = 0.80, 0.90$. The Asymptotic Power values for T_{0n} were calculated for the combined Sample Sizes values $N_t = n_1 + n_2$ with $\rho_1 = 1$ from (65) + (66). Relative Efficiency values were approximated using (69). A Relative Efficiency value less (or greater) than one implies a larger (or smaller) Asymptotic Power value for the T_{0n} test versus the Asymptotic Power value for the $\tilde{\mathrm{Z}}_n$ test.

Table 3: Power simulation results for the $\tilde{Z}_n$ and T_{0n} tests, using 500 independent runs, where $X_1 \sim \text{Gamma}(1, \beta_{\gamma 1})$, $X_2 \sim \text{Gamma}(1, 3)$.

$\beta_{\gamma 1}$	α_{H0}	Sample Sizes n_1, n_2	Sample Levels $\tilde{Z}_n$, T_{0n}	Sample Powers $\tilde{Z}_n$, T_{0n}	Asymptotic Powers $\tilde{Z}_n(\gamma)$, T_{0n}	Rel. Eff. (69)
2	.05	84, 84	.050, .056	.776, .696	.8010, .7344	1.1662
2	.05	122, 122	.058, .050	.926, .904	.9002, .8825	1.0584
2	.01	119, 119	.018, .012	.822, .698	.8008, .6858	1.2345
2	.01	164, 164	.016, .006	.892, .838	.9009, .8532	1.1285
2.5	.05	442, 442	.044, .042	.786, .748	.8002, .7715	1.0733
2.5	.05	614, 614	.056, .052	.908, .878	.9004, .8911	1.0321
2.5	.01	644, 644	.014, .010	.798, .754	.8001, .7531	1.0978
2.5	.01	848, 848	.010, .012	.908, .880	.9001, .8793	1.0592
3.5	.05	707, 707	.048, .046	.792, .830	.8005, .8251	0.9381
3.5	.05	920, 920	.054, .052	.894, .910	.9001, .9180	0.9623
3.5	.01	1068, 1068	.014, .010	.790, .836	.8003, .8365	0.9247
3.5	.01	1327, 1327	.010, .006	.882, .908	.9001, .9180	0.9461
4	.05	217, 217	.042, .030	.802, .858	.8004, .8475	0.8843
4	.05	277, 277	.058, .054	.874, .906	.9007, .9227	0.9217
4	.01	332, 332	.010, .010	.802, .890	.8008, .8667	0.8641
4	.01	405, 405	.012, .010	.896, .936	.9006, .9341	0.8965

In Table 3, the important columns to compare are the Sample and Asymptotic Powers columns. The Sample Powers values for the $\tilde{Z}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level. The Sample Levels values for the $\tilde{Z}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level when the null hypothesis was true. For these simulations in Table 3, the Sample Sizes values are large enough so that the Sample Powers values are in agreement with the corresponding Asymptotic Powers values. Also for these simulations, the Sample Powers values for the $\tilde{Z}_n$ and T_{0n} tests support the Relative Efficiency values. A larger (or smaller) Sample Power value for the T_{0n} test versus the Sample Power value for the $\tilde{Z}_n$ test is compatible with the smaller (or larger) than one Relative Efficiency value.

2.2.3.3 Gamma Example II As another example, assume X_1 and X_2 have Gamma distributions with different shape parameters $\alpha_{\gamma 1} \neq \alpha_{\gamma 2}$ and with a common scale parameter β_γ as described in section 2.1.1.3. Note that $h(x) = \log(x)$. For this Gamma Example II, the coefficients in the relative efficiency equation (71) are specialized to the coefficients in (76) and (77) below

$$\begin{aligned}
x_{11}, \ldots, x_{1n_1} &\sim X_1 \text{ with Gamma}\left(\alpha_{\gamma 1}, \beta_\gamma\right) \text{ pdf} \\
x_{21}, \ldots, x_{2n_2} &\sim X_2 \text{ with Gamma}\left(\alpha_{\gamma 2}, \beta_\gamma\right) \text{ pdf} \\
\left(\mu_j, \sigma_j^2\right) &= \left(\alpha_{\gamma j}\beta_\gamma,\ \alpha_{\gamma j}\beta_\gamma^2\right),\ j = 1, 2 \\
\left(\alpha_0, \beta_0\right) &= \left(\log\frac{\Gamma(\alpha_{\gamma 2})}{\Gamma(\alpha_{\gamma 2})} + \left(\alpha_{\gamma 2} - \alpha_{\gamma 1}\right)\log\beta_\gamma,\ \left(\alpha_{\gamma 1} - \alpha_{\gamma 2}\right)\right) \\
\sigma_h^2 &= \frac{\Gamma''\left(\alpha_{\gamma 2}\right)}{\Gamma\left(\alpha_{\gamma 2}\right)} - \left(\frac{\Gamma'\left(\alpha_{\gamma 2}\right)}{\Gamma\left(\alpha_{\gamma 2}\right)}\right)^2 \\
\frac{\sigma_p}{\left(\mu_1 - \mu_2\right)} &= \sqrt{\frac{\rho_1\alpha_{\gamma 1} + \alpha_{\gamma 2}}{\rho_1 + 1}} / \left(\alpha_{\gamma 1} - \alpha_{\gamma 2}\right) && (76) \\
\sigma_h\beta_0\frac{\sigma_p}{\left(\mu_1 - \mu_2\right)} &= \sigma_h\sqrt{\frac{\rho_1\alpha_{\gamma 1} + \alpha_{\gamma 2}}{\rho_1 + 1}}\ . && (77)
\end{aligned}$$

The asymptotic relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ follows directly. With regard to the Pitman efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$, the variance of T_0^* converges to one as $\alpha_{\gamma 1}$ approaches $\alpha_{\gamma 2}$ in (61). In general, as previously shown in (45), the variance of $\tilde{\mathrm{Z}}^*$ converges to one as β_0 approaches zero.

$$\begin{aligned}
\text{A.R.E } &\equiv \lim_{\gamma \to 1}\frac{\mathrm{N}_t}{\mathrm{N}_z} = \sigma_h^2\left(\frac{\rho_1\alpha_{\gamma 1} + \alpha_{\gamma 2}}{\rho_1 + 1}\right)\frac{\sigma^2\left(\mathrm{T}_0^*\right)}{\sigma^2\left(\tilde{\mathrm{Z}}^*\right)} \\
\lim_{\beta_0 \to 0}\sigma^2\left(\tilde{\mathrm{Z}}^*\right) &= 1, \quad \lim_{\left(\alpha_{\gamma 1} - \alpha_{\gamma 2}\right) \to 0}\sigma^2\left(\mathrm{T}_0^*\right) = 1 \\
e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) &\equiv \lim_{\sqrt{\mathrm{N}_z}\left(\alpha_{\gamma 1} - \alpha_{\gamma 2}\right) \to c}\frac{\mathrm{N}_t}{\mathrm{N}_z} = \sigma_h^2\alpha_{\gamma 2}
\end{aligned}$$

By inspection, σ_h^2 depends only on $\alpha_{\gamma 2}$, not on $\beta_{\gamma 2}$. Figures 3 and 7 graph the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* separately when $\beta_\gamma = 1$. Figure 15 graphs the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* together when $\beta_\gamma = 1$. Figure 16 graphs the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ when $\beta_\gamma = 1$, and when $\alpha_{H0} = .05$.

In Figure 16, the relative efficiency is greater than one, in a large neighborhood of $\alpha_{\gamma 1} = \alpha_{\gamma 2} = 3$. For smaller power values $\gamma = .7, .8, .9$, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ is greater than one when $\alpha_{\gamma 1} < \alpha_{\gamma 2} = 3$, except when $\alpha_{\gamma 1}$ is close to 1. As the power value increases, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ increases, so that at a power value of $\gamma = .9999$, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ is greater than one for $\alpha_{\gamma 1} = 1$. In contrast, the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ decreases for $\alpha_{\gamma 1} > \alpha_{\gamma 2} = 3$ as the power value increases.

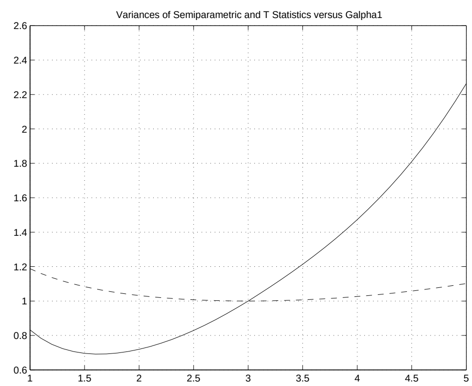

Figure 15: Given $X_1 \sim \mathrm{Gamma}(\alpha_{\gamma 1}, 1)$, $X_2 \sim \mathrm{Gamma}(3, 1)$, solid line is the variance of $\tilde{\mathrm{Z}}^*$ versus $\alpha_{\gamma 1}$, dashed line is the variance of T_0^* versus $\alpha_{\gamma 1}$.

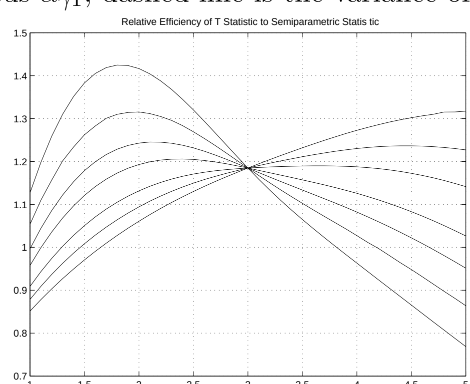

Figure 16: Relative Efficiency N_t/N_z curves of T_{0n} to $\tilde{\mathrm{Z}}_n$, versus $\alpha_{\gamma 1}$, when $X_1 \sim \mathrm{Gamma}(\alpha_{\gamma 1}, 1)$, $X_2 \sim \mathrm{Gamma}(3, 1)$, and when $\alpha_{H0} = .05$. The curves, starting from the bottom left, correspond to different power values of $\gamma = .7$, .8, .9, .99, .9999, .9999999999999999, 1.

The Pitman efficiency calculated using the slope formula (72) is consistent with the previous calculation. Let $\alpha_{\gamma 1n} \equiv \alpha_{\gamma 2} + \theta_n$.

$$\mu_T(\theta_n) = \frac{\mu_1(\theta_n) - \mu_2}{\sigma_p(\theta_n)} = \frac{\theta_n}{\alpha_{\gamma 2} + \frac{1}{1+\rho_1}\theta_n},\ \mu_T'(0) = \alpha_{\gamma 2}^{-\frac{1}{2}}$$

$$\mu_Z(\theta_n) = \sigma_h \beta_0(\theta_n) = \sigma_h \theta_n,\ \mu_Z'(0) = \sigma_h$$

$$e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) = \left(\frac{\mu_Z'(0)}{\mu_T'(0)}\right)^2 = \sigma_h^2 \alpha_{\gamma 2}$$

Figure 17 graphs the Pitman efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ over a range of $\alpha_{\gamma 2}$. This figure shows that the Pitman efficiency decreases towards one as a function of $\alpha_{\gamma 2}$.

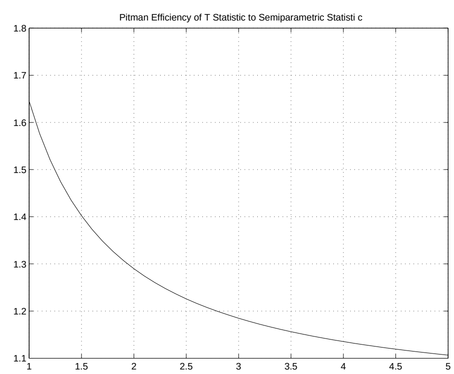

Figure 17: Pitman Efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$, versus $\alpha_{\gamma 2}$, when $X_2 \sim$ Gamma$(\alpha_{\gamma 2}, 1)$.

Power simulation results for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests in Table 4 show how well the asymptotic power approximates finite sample behavior where $X_1 \sim$ Gamma$(\alpha_{\gamma 1}, 1)$, $X_2 \sim$ Gamma$(3, 1)$, and where $\alpha_{\gamma 1} = 2, 2.5, 3.5, 4$. The combined Sample Sizes values $N_z = n_1 + n_2$ were calculated with $\rho_1 = 1$ via (67) and (63) + (64) to provide the specified $\alpha_{H0} = 0.05, 0.01$ error and to provide the specified Asymptotic Power values for $\tilde{\mathrm{Z}}_n$ that approximate a power of $\gamma = 0.80, 0.90$. The Asymptotic Power values for T_{0n} were calculated for the combined Sample Sizes values $N_t = n_1 + n_2$ with $\rho_1 = 1$ from (65) + (66). Relative Efficiency values were approximated using (69). A Relative Efficiency value greater than one implies a smaller Asymptotic Power value for the T_{0n} test versus the Asymptotic Power value for the $\tilde{\mathrm{Z}}_n$ test.

Table 4: Power simulation results for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests, using 500 independent runs, where $X_1 \sim \mathrm{Gamma}(\alpha_{\gamma 1}, 1)$, $X_2 \sim \mathrm{Gamma}(3, 1)$.

$\alpha_{\gamma 1}$	α_{H0}	Sample Sizes n_1, n_2	Sample Levels $\tilde{\mathrm{Z}}_n$, T_{0n}	Sample Powers $\tilde{\mathrm{Z}}_n$, T_{0n}	Asymptotic Powers $\tilde{\mathrm{Z}}_n(\gamma)$, T_{0n}	Rel. Eff. (69)
2	.05	37, 37	.056, .050	.812, .770	.8095, .7729	1.0944
2	.05	48, 48	.056, .054	.906, .878	.9064, .8688	1.1316
2	.01	55, 55	.010, .008	.814, .772	.8020, .7671	1.0739
2	.01	68, 68	.012, .014	.882, .848	.9003, .8632	1.1066
2.5	.05	151, 151	.054, .056	.776, .740	.8013, .7446	1.1496
2.5	.05	199, 199	.044, .042	.886, .848	.9015, .8517	1.1708
2.5	.01	227, 227	.006, .010	.816, .746	.8017, .7370	1.1377
2.5	.01	284, 284	.008, .012	.876, .830	.9004, .8446	1.1566
3.5	.05	169, 169	.050, .048	.814, .706	.8004, .7216	1.2111
3.5	.05	231, 231	.040, .046	.916, .856	.9009, .8455	1.1896
3.5	.01	249, 249	.012, .008	.814, .704	.8008, .6975	1.2237
3.5	.01	323, 323	.010, .010	.890, .824	.9009, .8278	1.2038
4	.05	46, 46	.046, .042	.806, .722	.8075, .7243	1.2302
4	.05	63, 63	.058, .054	.890, .836	.9017, .8477	1.1876
4	.01	66, 66	.014, .012	.816, .712	.8030, .6874	1.2556
4	.01	87, 87	.008, .012	.914, .838	.9020, .8257	1.2157

In Table 4, the important columns to compare are the Sample and Asymptotic Powers columns. The Sample Powers values for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level. The Sample Levels values for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level when the null hypothesis was true. For these simulations in Table 4, the Sample Sizes values are large enough so that the Sample Power values are in agreement with the corresponding Asymptotic Power values. Also for these simulations, the Sample Powers values for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests support the Relative Efficiency values. A smaller Sample Power value for the T_{0n} test versus the Sample Power value for the $\tilde{\mathrm{Z}}_n$ test is compatible with the larger than one Relative Efficiency value.

2.2.3.4 Log Normal Example As another example, assume X_1 and X_2 have log normal distributions with different $\mu_{l1} \neq \mu_{l2}$ parameters and with a common σ_l^2 parameter as described in section 2.1.1.4. Note that $h(x) = \log(x)$. For this log normal example, the coefficients in the relative efficiency equation (71) are specialized to the coefficients in (78) and (79) below

$$
\begin{aligned}
x_{11}, \ldots, x_{1n_1} &\sim X_1 \text{ with LN}\left(\mu_{l1}, \sigma_l^2\right) \text{ pdf} \\
x_{21}, \ldots, x_{2n_2} &\sim X_2 \text{ with LN}\left(\mu_{l2}, \sigma_l^2\right) \text{ pdf} \\
\left(\mu_j, \sigma_j^2\right) &= \left(e^{\mu_{lj}+\sigma_l^2/2},\ e^{2\mu_{lj}+\sigma_l^2}\left(e^{\sigma_l^2}-1\right)\right),\ j = 1, 2 \\
\left(\alpha_0, \beta_0\right) &= \left(\frac{\mu_{l2}^2-\mu_{l1}^2}{2\sigma_l^2},\ \frac{\mu_{l1}-\mu_{l2}}{\sigma_l^2}\right),\ \sigma_h^2 = \sigma_l^2 \\
\frac{\sigma_p}{\left(\mu_1-\mu_2\right)} &= \sqrt{\frac{\left(\rho_1 e^{2\left(\mu_{l1}-\mu_{l2}\right)}+1\right)\left(e^{\sigma_l^2}-1\right)}{\rho_1+1}}\left(\frac{1}{e^{\mu_{l1}-\mu_{l2}}-1}\right) \qquad (78) \\
\sigma_h \beta_0 \frac{\sigma_p}{\left(\mu_1-\mu_2\right)} &= \frac{1}{\sigma_l}\sqrt{\frac{\left(\rho_1 e^{2\left(\mu_{l1}-\mu_{l2}\right)}+1\right)\left(e^{\sigma_l^2}-1\right)}{\rho_1+1}}\left(\frac{\mu_{l1}-\mu_{l2}}{e^{\mu_{l1}-\mu_{l2}}-1}\right). \qquad (79)
\end{aligned}
$$

The asymptotic relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ follows directly. With regard to the Pitman efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$, as previously shown in (62), the variance of T_0^* converges to one, since $\left(\mu_1, \sigma_1^2\right)$ approaches $\left(\mu_2, \sigma_2^2\right)$ as μ_{l1} approaches μ_{l2}. Also as previously shown in (45), the variance of $\tilde{\mathrm{Z}}^*$ converges to one as β_0 approaches zero.

$$
\begin{aligned}
\text{A.R.E.} \equiv \lim_{\gamma \to 1} \frac{\mathrm{N}_t}{\mathrm{N}_z} &= \left(\sigma_h \beta_0 \frac{\sigma_p}{\left(\mu_1-\mu_2\right)}\right)^2 \frac{\sigma^2\left(\mathrm{T}_0^*\right)}{\sigma^2\left(\tilde{\mathrm{Z}}^*\right)} \\
\lim_{\beta_0 \to 0} \sigma^2\left(\tilde{\mathrm{Z}}^*\right) = 1, &\quad \lim_{\left(\mu_{l1}-\mu_{l2}\right) \to 0} \sigma^2\left(\mathrm{T}_0^*\right) = 1 \\
e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) \equiv \lim_{\sqrt{\mathrm{N}_z}\left(\mu_{l1}-\mu_{l2}\right) \to c} \frac{\mathrm{N}_t}{\mathrm{N}_z} &= \frac{1}{\sigma_l^2}\left(e^{\sigma_l^2}-1\right)
\end{aligned}
$$

Figures 4 and 8 graph the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* separately when $\sigma_l^2 = 1$. Figure 18 graphs the variances of $\tilde{\mathrm{Z}}^*$ and T_0^* together when $\sigma_l^2 = 1$. Figure 19 graphs the relative efficiency of T_{0n} to $\tilde{\mathrm{Z}}_n$ when $\sigma_l^2 = 1$, and when $\alpha_{H0} = .05$.

In Figure 19, the relative efficiency is greater than one, for $\mu_{l1} \in (-2, 2)$. In fact, the relative efficiency increases as the power value increases, or as the difference $|\mu_{l1} - \mu_{l2}|$ increases.

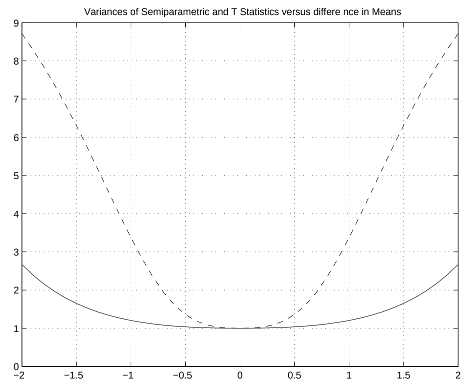

Figure 18: Given $X_1 \sim \text{LN}(\mu_{l1}, 1)$, $X_2 \sim \text{LN}(0, 1)$, the solid line is the variance of $\tilde{\text{Z}}^*$ versus μ_{l1}, the dashed line is the variance of T_0^* versus μ_{l1}.

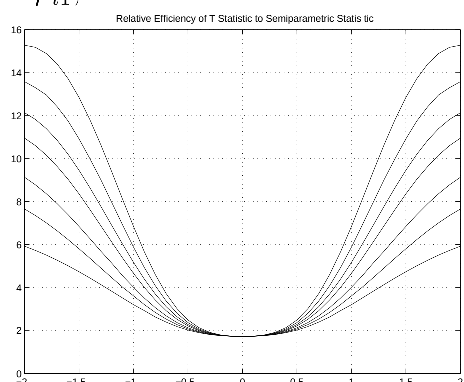

Figure 19: Relative Efficiency N_t/N_z curves of T_{0n} to $\tilde{\text{Z}}_n$, versus μ_{l1}, when $X_1 \sim \text{LN}(\mu_{l1}, 1)$, $X_2 \sim \text{LN}(0, 1)$, and when $\alpha_{H0} = .05$. The curves, starting from the bottom left, correspond to different power values of $\gamma = .7$, .8, .9, .99, .9999, .9999999999999999, 1.

The Pitman efficiency calculated using the slope formula (72) is consistent with the previous calculation. Let $\mu_{l1n} \equiv \mu_{l2} + \theta_n$.

$$\mu_T(\theta_n) = \frac{\mu_1(\theta_n) - \mu_2}{\sigma_p(\theta_n)} = \frac{e^{\theta_n} - 1}{\left(e^{\sigma_l^2} - 1\right)^{\frac{1}{2}} \left(\frac{\rho_1}{1+\rho_1} e^{2\theta_n} + \frac{1}{1+\rho_1}\right)^{\frac{1}{2}}}$$

$$\mu_T'(0) = \left(e^{\sigma_l^2} - 1\right)^{-\frac{1}{2}}$$

$$\mu_Z(\theta_n) = \sigma_h \beta_0(\theta_n) = \frac{\theta_n}{\sigma_l}, \ \mu_Z'(0) = \frac{1}{\sigma_l}$$

$$e\left(\mathrm{T}_{0n}, \tilde{\mathrm{Z}}_n\right) = \left(\frac{\mu_Z'(0)}{\mu_T'(0)}\right)^2 = \frac{e^{\sigma_l^2} - 1}{\sigma_l^2}$$

Power simulation results for the $\tilde{\mathrm{Z}}_n$ and T_{0n} tests in Table 5 show how well the asymptotic power approximates finite sample behavior where $X_1 \sim \mathrm{LN}(\mu_{l1}, 1)$, $X_2 \sim \mathrm{LN}(0, 1)$, and where $\mu_{l1} = .2, .3, .4, .5$. The combined Sample Sizes values $N_t = n_1 + n_2$ were calculated with $\rho_1 = 1$ via (68) and (65) + (66) to provide the specified $\alpha_{H0} = 0.05, 0.01$ error and to provide the specified Asymptotic Power values for T_{0n} that approximates a power of $\gamma = 0.80, 0.90$. The Asymptotic Power values for $\tilde{\mathrm{Z}}_n$ were calculated for the combined sample size $N_z = n_1 + n_2$ with $\rho_1 = 1$ from (63) + (64). Relative Efficiency values were approximated using (69). A Relative Efficiency value greater than one implies a smaller Asymptotic Power value for the T_{0n} test versus the Asymptotic Power value for the $\tilde{\mathrm{Z}}_n$ test.

Table 5: Power simulation results for the $\tilde{Z}_n$ and T_{0n} tests, using 500 independent runs, where $X_1 \sim LN(\mu_{l1}, 1)$, $X_2 \sim LN(0, 1)$.

μ_{l1}	α_{H0}	Sample Sizes n_1, n_2	Sample Levels $\tilde{Z}_n$, T_{0n}	Sample Powers $\tilde{Z}_n$, T_{0n}	Asymptotic Powers $\tilde{Z}_n$, $T_{0n}(\gamma)$	Rel. Eff. (69)
.2	.05	692, 692	.046, .050	.950, .820	.9604, .8002	1.7597
.2	.05	929, 929	.048, .044	.989, .926	.9905, .9002	1.7637
.2	.01	1028, 1028	.008, .012	.972, .806	.9746, .8002	1.7574
.2	.01	1313, 1313	.008, .010	.996, .916	.9945, .9002	1.7610
.3	.05	319, 319	.050, .050	.974, .836	.9655, .8004	1.8206
.3	.05	430, 430	.046, .044	.990, .912	.9923, .9000	1.8326
.3	.01	473, 473	.010, .010	.966, .836	.9786, .8009	1.8138
.3	.01	606, 606	.006, .010	.996, .926	.9957, .9002	1.8246
.4	.05	190, 190	.048, .056	.970, .858	.9724, .8006	1.9204
.4	.05	259, 259	.054, .052	.998, .932	.9948, .9009	1.9483
.4	.01	280, 280	.008, .014	.976, .856	.9836, .8007	1.9045
.4	.01	362, 362	.010, .010	.998, .938	.9972, .9004	1.9297
.5	.05	132, 132	.056, .046	.970, .868	.9805, .8019	2.0680
.5	.05	182, 182	.048, .054	.994, .944	.9971, .9014	2.1230
.5	.01	193, 193	.010, .008	.992, .868	.9891, .8022	2.0369
.5	.01	252, 252	.012, .012	.996, .960	.9986, .9008	2.0864

In Table 5, the important columns to compare are the Sample and Asymptotic Powers columns. The Sample Powers values for the $\tilde{Z}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level. The Sample Levels values for the $\tilde{Z}_n$ and T_{0n} tests identify the proportion of simulation runs that failed the $\mathbf{H}_0$ test at the α_{H0} level when the null hypothesis is true. For these simulations, the Sample Sizes values are large enough so that the Sample Power values for the $\tilde{Z}_n$ test are in agreement with the corresponding Asymptotic Power values. For these simulations, the Sample Sizes values are not large enough in general so that the Sample Power values for the T_{0n} test are not in agreement in general with the corresponding Asymptotic Power values. Also for these simulations, the Sample Powers values for the $\tilde{Z}_n$ and T_{0n} tests support the Relative Efficiency values. A smaller Sample Power value for the T_{0n} test versus the Sample Power value for the $\tilde{Z}_n$ test is compatible with the larger than one Relative Efficiency value.

3 Computational Aspects of State Space Models

This section develops an asymptotic theory for state space smoother precisions and introduces a partial state space smoother. Subsection 3.1 defines a general multivariate linear Gaussian state space model and provides several examples of an ARMA time series that is recast in terms of a linear Gaussian state space model. Subsection 3.2 identifies and shows the formulas for the Kalman Predictor, Filter, and Smoother. Subsection 3.3 develops a likelihood smoother form of the state space smoother based on the general multivariate version of the linear Gaussian state space model introduced in subsection 3.1. Subsection 3.4 applies the likelihood smoother to a univariate version of the linear Gaussian state space model with constant parameters in order to develop various bounds on the smoother precisions, to develop simple formulas for the smoother estimates and precisions, and to develop limits for the smoother precisions. Subsection 3.4.1 generalizes this theory to account for missing observations. Subsection 3.5 introduces the concept of a partial state space smoother and provides several examples.

3.1 Linear Gaussian State Space Models

This section on linear Gaussian state space models is adopted from Kedem and Fokianos (2002) [14]. Let $\boldsymbol{\beta}_{0:N} = \{\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_N\}$ represent a sequence of $N+1$ (unknown) states, $\boldsymbol{\mathcal{F}}_N = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$ a sequence of N observations, and $\boldsymbol{\mathcal{X}}_N = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$ the corresponding covariate sequence. Let $\boldsymbol{\mathcal{F}}_t$ represent the information available to the observer at time t using the following convention:

$$\boldsymbol{\mathcal{F}}_0 = \boldsymbol{\emptyset},\ \boldsymbol{\mathcal{F}}_t = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{t-1},\ \boldsymbol{Y}_t\} = \{\boldsymbol{\mathcal{F}}_{t-1}, \boldsymbol{Y}_t\}\ .$$

The linear Gaussian state space model is defined by the following linear system of equations:

$$\begin{array}{lll} \text{Initial Information:} & \boldsymbol{\beta}_0 \sim \mathrm{N}_p(\boldsymbol{b}_0, \boldsymbol{W}_0) & \\ \text{System Equation:} & \boldsymbol{\beta}_t = \boldsymbol{F}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t \sim \mathrm{N}_p(\boldsymbol{0}, \boldsymbol{W}_t) \\ \text{Observation Equation:} & \boldsymbol{Y}_t = \boldsymbol{z}_t'\boldsymbol{\beta}_t + \boldsymbol{v}_t, & \boldsymbol{v}_t \sim \mathrm{N}_q(\boldsymbol{0}, \boldsymbol{V}_t) \end{array} \tag{80}$$

where $\{\boldsymbol{\beta}_0\}$, $\{\boldsymbol{w}_t : t = 1, \ldots, N\}$, and $\{\boldsymbol{v}_t : t = 1, \ldots, N\}$ are mutually independent collections of independent random vectors; where the system equation is true for $t = 1, \ldots, N$ and the observation equation is true for all $\boldsymbol{Y}_t \in \boldsymbol{\mathcal{F}}_N$, i.e. for $t = 1, \ldots, N$; where all distribution parameters $\{\boldsymbol{b}_0, \boldsymbol{W}_0, \boldsymbol{W}_t, \boldsymbol{V}_t$ for $t = 1, \ldots, N\}$ are known; where $\boldsymbol{F}_t$ for $t = 1, \ldots, N$

are known matrices; and where $\boldsymbol{z}_t$ for $t = 1, \ldots, N$ are known matrices that may contain covariates from $\boldsymbol{\mathcal{X}}_t$ such as past observations or may contain parameters that are known at time t. Each state $\boldsymbol{\beta}_t$ for $t = 0, \ldots, N$ can be thought of as an unknown covariate or as an unknown random parameter at time t. Thus the concept of "state" in the linear Gaussian state space model can be interpreted in several ways.

3.1.1 Examples of Linear State Space Models

An ARMA(p, q) process defined by $\phi(B)Y_t = \theta(B)w_t$ where:

$$\begin{aligned} BY_t &= Y_{t-1}, \\ \phi(B) &= 1 - \phi_1 B - \cdots - \phi_p B^p, \\ \theta(B) &= 1 + \theta_1 B + \cdots + \theta_q B^q, \end{aligned}$$

has many state space representations (80). Kedem and Fokianos (2002) [14] developed one such representation for the ARMA(p, q) process by using:

$$\begin{aligned} \phi(B)X_t = w_t \text{ or } X_t &= \phi^{-1}(B)w_t, \\ Y_t &= \theta(B)X_t = \theta(B)\phi^{-1}(B)w_t, \\ \phi(B)Y_t &= \theta(B)w_t\ . \end{aligned}$$

The corresponding state space model can be written as:

$$\begin{aligned} \boldsymbol{\beta}_t &= \begin{pmatrix} \phi_1 & \cdots & \phi_{r-2} & \phi_{r-1} & \phi_r \\ 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \boldsymbol{\beta}_{t-1} + \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} w_t \\ \boldsymbol{\beta}_t &= (X_t, \ldots, X_{t-r+1})' \\ Y_t &= \begin{pmatrix} 1 & \theta_1 & \theta_2 & \cdots & \theta_{r-1} \end{pmatrix} \boldsymbol{\beta}_t \end{aligned}$$

where $r = \max(p, q+1)$, where $\phi_j = 0$ for $j > p$, and where $\theta_j = 0$ for $j > q$.

Durbin and Koopman (2001) [7] provide an alternate state space representation for the ARMA(p, q) process as follows

$$\boldsymbol{\beta}_t = \begin{bmatrix} \phi_1 & 1 & & 0 \\ \vdots & & \ddots & \\ \phi_{r-1} & 0 & & 1 \\ \phi_r & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{\beta}_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{pmatrix} w_t$$

$$\boldsymbol{\beta}_t = \begin{pmatrix} Y_t \\ \phi_2 Y_{t-1} + \cdots + \phi_r Y_{t-r+1} + \theta_1 w_t + \cdots + \theta_{r-1} w_{t-r+2} \\ \phi_3 Y_{t-1} + \cdots + \phi_r Y_{t-r+2} + \theta_2 w_t + \cdots + \theta_{r-1} w_{t-r+3} \\ \vdots \\ \theta_r Y_{t-1} + \theta_{r-1} w_t \end{pmatrix}$$

$$Y_t = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \end{pmatrix} \boldsymbol{\beta}_t \ .$$

Durbin and Koopman [7] also provide a state space representation for the ARIMA(p, d, q) process as defined by $\phi(B)(1 - B)^d Y_t = \theta(B) w_t$.

3.2 Kalman Predictor/Filter and State Space Smoother

Given a sequence of observations $\boldsymbol{\mathcal{F}}_N = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$, the linear state space model is used to estimate the (unknown) state sequence $\boldsymbol{\beta}_{0:t} = \{\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_t\}$. The estimation of $\boldsymbol{\beta}_t$ given $\boldsymbol{\mathcal{F}}_s$, or the estimation of its conditional distribution $f(\boldsymbol{\beta}_t|\boldsymbol{\mathcal{F}}_s)$, $s \leq N$, is called prediction if $t > s$; filtering if $t = s$; or smoothing if $t < s$.

In the Gaussian case of the linear state space model, the Kalman Prediction and Filtering methods and the Space Space Smoothing method calculate the conditional mean vector and the precision matrix of $\boldsymbol{\beta}_t|\boldsymbol{\mathcal{F}}_s$. For $t = 1, \ldots, N$ let

$$\boldsymbol{\beta}_{t|s} = \mathrm{E}[\boldsymbol{\beta}_t|\boldsymbol{\mathcal{F}}_s], \ \boldsymbol{P}_{t|s} = \mathrm{E}[(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|s})(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|s})'] \ .$$

The covariance matrix, between the residuals $\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|s}$ and the observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_s$, being zero for all t and s implies that $\boldsymbol{P}_{t|s}$ is also the conditional variance of $\boldsymbol{\beta}_t|\boldsymbol{\mathcal{F}}_s$, i.e.

$$\boldsymbol{P}_{t|s} = \mathrm{E}[(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|s})(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|s})'] = \mathrm{E}[(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|s})(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|s})'|\boldsymbol{\mathcal{F}}_s] = \mathrm{Var}(\boldsymbol{\beta}_t|\boldsymbol{\mathcal{F}}_s).$$

Letting $\boldsymbol{\beta}_{0|0} = \boldsymbol{b}_0, \boldsymbol{P}_{0|0} = \boldsymbol{W}_0$, and using the initial condition $\boldsymbol{\beta}_0|\boldsymbol{\mathcal{F}}_0 \sim \mathrm{N}_p(\boldsymbol{\beta}_{0|0}, \boldsymbol{P}_{0|0})$, leads to the following Kalman methods, see [14].

The Kalman Prediction method, for $t = 1 \ldots N$, calculates:

$$\begin{aligned} \boldsymbol{\beta}_{t|t-1} &= \boldsymbol{F}_t \boldsymbol{\beta}_{t-1|t-1}, \\ \boldsymbol{P}_{t|t-1} &= \boldsymbol{F}_t \boldsymbol{P}_{t-1|t-1} \boldsymbol{F}_t' + \boldsymbol{W}_t. \end{aligned}$$

The Kalman Filtering method, for $t = 1 \dots N$, where $\boldsymbol{K}_t$ is the Kalman Gain, calculates:

$$\begin{aligned}
\boldsymbol{\beta}_{t|t} &= \boldsymbol{\beta}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \boldsymbol{z}_t'\boldsymbol{\beta}_{t|t-1}),\\
\boldsymbol{P}_{t|t} &= [\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{z}_t']\boldsymbol{P}_{t|t-1},\\
\boldsymbol{K}_t &\equiv \boldsymbol{P}_{t|t-1}\boldsymbol{z}_t[\boldsymbol{z}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{z}_t + \boldsymbol{V}_t]^{-1}.
\end{aligned}$$

The State Space Smoothing method, for $t = N \dots 1$, calculates:

$$\begin{aligned}
\boldsymbol{\beta}_{t-1|N} &= \boldsymbol{\beta}_{t-1|t-1} + \boldsymbol{B}_t(\boldsymbol{\beta}_{t|N} - \boldsymbol{\beta}_{t|t-1}),\\
\boldsymbol{P}_{t-1|N} &= \boldsymbol{P}_{t-1|t-1} + \boldsymbol{B}_t(\boldsymbol{P}_{t|N} - \boldsymbol{P}_{t|t-1})\boldsymbol{B}_t',\\
\boldsymbol{B}_t &\equiv \boldsymbol{P}_{t-1|t-1}\boldsymbol{F}_t'\boldsymbol{P}_{t|t-1}^{-1}.
\end{aligned}$$

The Kalman Prediction result follows immediately from using the State Space equations (80) given $\boldsymbol{\beta}_{t-1}|\boldsymbol{\mathcal{F}}_{t-1} \sim \mathrm{N}_p(\boldsymbol{\beta}_{t-1|t-1}, \boldsymbol{P}_{t-1|t-1})$.

The Kalman Filtering result follows from using the State Space equations (80) and the Kalman Prediction result to show:

$$\left.\begin{pmatrix}\boldsymbol{\beta}_t\\ \boldsymbol{Y}_t\end{pmatrix}\right|\boldsymbol{\mathcal{F}}_{t-1} \sim \mathrm{N}_{p+q}\left[\begin{pmatrix}\boldsymbol{\beta}_{t|t-1}\\ \boldsymbol{z}_t'\boldsymbol{\beta}_{t|t-1}\end{pmatrix}, \begin{pmatrix}\boldsymbol{P}_{t|t-1} & \boldsymbol{P}_{t|t-1}\boldsymbol{z}_t\\ \boldsymbol{z}_t'\boldsymbol{P}_{t|t-1} & \boldsymbol{z}_t'\boldsymbol{P}_{t|t-1}\boldsymbol{z}_t + V_t\end{pmatrix}\right]$$

and by applying the Normal distribution to Conditional Normal distribution transformation:

$$\begin{aligned}
\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{Y}\end{pmatrix} &\sim \mathrm{N}_{p+q}\left[\begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{\beta}}\\ \boldsymbol{\mu}_{\boldsymbol{Y}}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}} & \boldsymbol{\Sigma}_{\boldsymbol{\beta Y}}\\ \boldsymbol{\Sigma}_{\boldsymbol{Y\beta}} & \boldsymbol{\Sigma}_{\boldsymbol{YY}}\end{pmatrix}\right]\\
\boldsymbol{\beta}|\boldsymbol{Y} &\sim \mathrm{N}_p(\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{Y}})\\
\boldsymbol{\mu}_{\boldsymbol{\beta}|\boldsymbol{Y}} &= \mathrm{E}[\boldsymbol{\beta}|\boldsymbol{Y}] = \boldsymbol{\mu}_{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\boldsymbol{\beta Y}}\boldsymbol{\Sigma}_{\boldsymbol{YY}}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}})\\
\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\boldsymbol{Y}} &= \mathrm{Var}[\boldsymbol{\beta}|\boldsymbol{Y}] = \boldsymbol{\Sigma}_{\boldsymbol{\beta\beta}} - \boldsymbol{\Sigma}_{\boldsymbol{\beta Y}}\boldsymbol{\Sigma}_{\boldsymbol{YY}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{Y\beta}}\ .
\end{aligned}$$

Derivation of the State Space Smoothing result is lengthly using a classical statistical approach. A Bayesian approach, due to Künsch (2001) [17], follows. For $t \leq N-1$, consider

$$\begin{aligned}
f\left(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1}, \boldsymbol{\mathcal{F}}_N\right) = f\left(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1}, \boldsymbol{\mathcal{F}}_t\right) &= \frac{f\left(\boldsymbol{\beta}_{t+1}|\boldsymbol{\beta}_t\right) f\left(\boldsymbol{\beta}_t|\boldsymbol{\mathcal{F}}_t\right)}{f\left(\boldsymbol{\beta}_{t+1}|\boldsymbol{\mathcal{F}}_t\right)}\\
&\propto \exp\left[-(\boldsymbol{\beta}_t - \boldsymbol{F}_{t+1}^{-1}\boldsymbol{\beta}_{t+1})'\frac{\boldsymbol{F}_{t+1}'\boldsymbol{W}_{t+1}^{-1}\boldsymbol{F}_{t+1}}{2}(\boldsymbol{\beta}_t - \boldsymbol{F}_{t+1}^{-1}\boldsymbol{\beta}_{t+1})\right.\\
&\qquad \left. -(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t})'\frac{\boldsymbol{P}_{t|t}^{-1}}{2}(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|t})\right]
\end{aligned}$$

where the proportionality constant does not depend on $\boldsymbol{\beta}_t$. Completing the square in the previous display where $(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1}, \boldsymbol{\mathcal{F}}_N) \sim \mathrm{N}(\boldsymbol{m}_t, \boldsymbol{R}_t)$ and where

$$\begin{aligned}
\boldsymbol{R}_t^{-1} &= \boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}\boldsymbol{F}_{t+1} + \boldsymbol{P}_{t|t}^{-1} \\
\boldsymbol{m}_t &= \boldsymbol{R}_t\left[\boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}\boldsymbol{\beta}_{t+1} + \boldsymbol{P}_{t|t}^{-1}\boldsymbol{\beta}_{t|t}\right] \\
&= \boldsymbol{R}_t\left[\boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}\boldsymbol{\beta}_{t+1} + \boldsymbol{R}_t^{-1}\boldsymbol{\beta}_{t|t} - \boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}\boldsymbol{F}_{t+1}\boldsymbol{\beta}_{t|t}\right] \\
&= \boldsymbol{\beta}_{t|t} + \boldsymbol{R}_t\boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}(\boldsymbol{\beta}_{t+1} - \boldsymbol{\beta}_{t+1|t}).
\end{aligned}$$

and then manipulating the following identify

$$\begin{aligned}
\left(\boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}\boldsymbol{F}_{t+1} + \boldsymbol{P}_{t|t}^{-1}\right)\boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1} &= \boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}\left(\boldsymbol{W}_{t+1} + \boldsymbol{F}_{t+1}\boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\right) \\
\boldsymbol{R}_t^{-1}\boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1} &= \boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1}\boldsymbol{P}_{t+1|t} \\
\boldsymbol{R}_t\boldsymbol{F}'_{t+1}\boldsymbol{W}_{t+1}^{-1} &= \boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\boldsymbol{P}_{t+1|t}^{-1} \\
\boldsymbol{R}_t &= \boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\boldsymbol{P}_{t+1|t}^{-1}\boldsymbol{W}_{t+1}\boldsymbol{F}'^{-1}_{t+1} \\
&= \boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\boldsymbol{P}_{t+1|t}^{-1}\left(\boldsymbol{P}_{t+1|t} - \boldsymbol{F}_{t+1}\boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\right)\boldsymbol{F}'^{-1}_{t+1}
\end{aligned}$$

gives the following conditional mean and conditional variance

$$\begin{aligned}
\boldsymbol{m}_t &= \boldsymbol{\beta}_{t|t} + \boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\boldsymbol{P}_{t+1|t}^{-1}(\boldsymbol{\beta}_{t+1} - \boldsymbol{\beta}_{t+1|t}) \\
\boldsymbol{R}_t &= \boldsymbol{P}_{t|t} - \boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\boldsymbol{P}_{t+1|t}^{-1}\boldsymbol{F}_{t+1}\boldsymbol{P}_{t|t}\ .
\end{aligned}$$

Using conditional expectation leads to

$$\begin{aligned}
\boldsymbol{\beta}_{t|N} &= \mathrm{E}\left(\boldsymbol{\beta}_t|\boldsymbol{\mathcal{F}}_N\right) = \mathrm{E}\left(\mathrm{E}\left(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t+1}, \boldsymbol{\mathcal{F}}_N\right)|\boldsymbol{\mathcal{F}}_N\right) = \mathrm{E}\left(\boldsymbol{m}_t|\boldsymbol{\mathcal{F}}_N\right) \\
&= \boldsymbol{\beta}_{t|t} + \boldsymbol{B}_{t+1}(\boldsymbol{\beta}_{t+1|N} - \boldsymbol{\beta}_{t+1|t}) \\
\boldsymbol{B}_{t+1} &\equiv \boldsymbol{P}_{t|t}\boldsymbol{F}'_{t+1}\boldsymbol{P}_{t+1|t}^{-1}\ .
\end{aligned}$$

Similarly for the Precision matrix

$$\begin{aligned}
\boldsymbol{P}_{t|N} &= \mathrm{E}\left[(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|N})(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t|N})'|\boldsymbol{\mathcal{F}}_N\right] \\
&= \mathrm{E}\left[(\boldsymbol{\beta}_t - \boldsymbol{m}_t)(\boldsymbol{\beta}_t - \boldsymbol{m}_t)'|\boldsymbol{\mathcal{F}}_N\right] \\
&+ \mathrm{E}\left[(\boldsymbol{m}_t - \boldsymbol{\beta}_{t|N})(\boldsymbol{m}_t - \boldsymbol{\beta}_{t|N})'|\boldsymbol{\mathcal{F}}_N\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[(\boldsymbol{\beta}_t - \boldsymbol{m}_t)(\boldsymbol{\beta}_t - \boldsymbol{m}_t)'|\boldsymbol{\beta}_{t+1}, \boldsymbol{\mathcal{F}}_N\right]|\boldsymbol{\mathcal{F}}_N\right] \\
&+ \mathrm{E}\left[(\boldsymbol{m}_t - \boldsymbol{\beta}_{t|N})(\boldsymbol{m}_t - \boldsymbol{\beta}_{t|N})'|\boldsymbol{\mathcal{F}}_N\right] \\
&= \mathrm{E}\left(\boldsymbol{R}_t|\boldsymbol{\mathcal{F}}_N\right) + \mathrm{Var}\left(\boldsymbol{m}_t|\boldsymbol{\mathcal{F}}_N\right) \\
&= \boldsymbol{P}_{t|t} - \boldsymbol{B}_{t+1}\boldsymbol{P}_{t+1|t}\boldsymbol{B}'_{t+1} + \boldsymbol{B}_{t+1}\boldsymbol{P}_{t+1|N}\boldsymbol{B}'_{t+1} \\
&= \boldsymbol{P}_{t|t} - \boldsymbol{B}_{t+1}\left(\boldsymbol{P}_{t+1|t} - \boldsymbol{P}_{t+1|N}\right)\boldsymbol{B}'_{t+1}\ .
\end{aligned}$$

3.3 Likelihood Smoother

Finding the mode of the posterior distribution for $\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_N$ provides an alternative method of deriving the state space smoother. The posterior distribution for $\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_N$ is given by:

$$f\left(\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_N\right) = \left[\prod_{t=1}^{N} f\left(\boldsymbol{Y}_t|\boldsymbol{\beta}_t\right)\right]\left[\prod_{t=1}^{N} f\left(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}\right)\right] f\left(\boldsymbol{\beta}_0\right)/f\left(\boldsymbol{\mathcal{F}}_N\right) \ . \tag{81}$$

The posterior log-likelihood function, ignoring a constant that depends only on $\boldsymbol{\mathcal{F}}_N$, is given by:

$$\log f\left(\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_N\right) = \sum_{t=1}^{N} \log f\left(\boldsymbol{Y}_t|\boldsymbol{\beta}_t\right) + \sum_{t=1}^{N} \log f\left(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}\right) + \log f\left(\boldsymbol{\beta}_0\right) \ . \tag{82}$$

When each of the conditional distributions has a Gaussian distribution:

$$\begin{aligned} \boldsymbol{Y}_t|\boldsymbol{\beta}_t &\sim f_{\boldsymbol{v}_t}\left(\boldsymbol{Y}_t - \boldsymbol{z}_t'\boldsymbol{\beta}_t\right) & &= \mathrm{N}_q\left(\mathbf{0}, \boldsymbol{V}_t\right) \\ \boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1} &\sim f_{\boldsymbol{w}_t}\left(\boldsymbol{\beta}_t - \boldsymbol{F}_t\boldsymbol{\beta}_{t-1}\right) & &= \mathrm{N}_p\left(\mathbf{0}, \boldsymbol{W}_t\right) \\ \boldsymbol{\beta}_0 &\sim f_{\boldsymbol{w}_0}\left(\boldsymbol{\beta}_0 - \boldsymbol{b}_0\right) & &= \mathrm{N}_p\left(\mathbf{0}, \boldsymbol{W}_0\right) \end{aligned} \tag{83}$$

then the posterior log-likelihood, ignoring a constant that does not depend on $\boldsymbol{\beta}_{0:N}$, is given by:

$$\begin{aligned} \log f\left(\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_N\right) = &-\frac{1}{2}\sum_{t=1}^{N}\left(\boldsymbol{Y}_t - \boldsymbol{z}_t'\boldsymbol{\beta}_t\right)'\boldsymbol{V}_t^{-1}\left(\boldsymbol{Y}_t - \boldsymbol{z}_t'\boldsymbol{\beta}_t\right) \\ &-\frac{1}{2}\sum_{t=1}^{N}\left(\boldsymbol{\beta}_t - \boldsymbol{F}_t\boldsymbol{\beta}_{t-1}\right)'\boldsymbol{W}_t^{-1}\left(\boldsymbol{\beta}_t - \boldsymbol{F}_t\boldsymbol{\beta}_{t-1}\right) \\ &-\frac{1}{2}\left(\boldsymbol{\beta}_0 - \boldsymbol{b}_0\right)'\boldsymbol{W}_0^{-1}\left(\boldsymbol{\beta}_0 - \boldsymbol{b}_0\right) \ . \end{aligned}$$

Finding the mode $\hat{\boldsymbol{\beta}}_{0:N} = \{\hat{\boldsymbol{\beta}}_{0|N}, \ldots, \hat{\boldsymbol{\beta}}_{N|N}\}$ of the posterior log-likelihood by maximizing the posterior log-likelihood using

$$\begin{aligned} \mathbf{0}_{(N+1\times p)} &= \boldsymbol{\nabla}\log f\left(\boldsymbol{\beta}_{0:N}|\,\boldsymbol{\mathcal{F}}_N\right)|_{\hat{\boldsymbol{\beta}}_{0:N}} \\ \boldsymbol{\nabla} &\equiv \left(\frac{\partial}{\partial\boldsymbol{\beta}_0}, \ldots, \frac{\partial}{\partial\boldsymbol{\beta}_N}\right)' \end{aligned}$$

leads to the following system of state estimating equations:

$$\begin{aligned}
\mathbf{0}' &= \frac{\partial}{\partial \boldsymbol{\beta}_0} \log f\left(\boldsymbol{\beta}_{0:N} | \boldsymbol{\mathcal{F}}_N\right)\Big|_{\hat{\boldsymbol{\beta}}_{0:N}} \\
&= \left(\hat{\boldsymbol{\beta}}_{1|N} - \boldsymbol{F}_1 \hat{\boldsymbol{\beta}}_{0|N}\right)' \boldsymbol{W}_1^{-1} \boldsymbol{F}_1 - \left(\hat{\boldsymbol{\beta}}_{0|N} - \boldsymbol{b}_0\right)' \boldsymbol{W}_0^{-1} \\
\mathbf{0}' &= \frac{\partial}{\partial \boldsymbol{\beta}_t} \log f\left(\boldsymbol{\beta}_{0:N} | \boldsymbol{\mathcal{F}}_N\right)\Big|_{\hat{\boldsymbol{\beta}}_{0:N}} \quad \text{for } t = 1, \ldots, N-1 \qquad (84) \\
&= \left(\hat{\boldsymbol{\beta}}_{t+1|N} - \boldsymbol{F}_{t+1} \hat{\boldsymbol{\beta}}_{t|N}\right)' \boldsymbol{W}_{t+1}^{-1} \boldsymbol{F}_{t+1} - \left(\hat{\boldsymbol{\beta}}_{t|N} - \boldsymbol{F}_t \hat{\boldsymbol{\beta}}_{t-1|N}\right)' \boldsymbol{W}_t^{-1} \\
&+ \left(\boldsymbol{Y}_t - \boldsymbol{z}_t' \hat{\boldsymbol{\beta}}_{t|N}\right)' \boldsymbol{V}_t^{-1} \boldsymbol{z}_t' \\
\mathbf{0}' &= \frac{\partial}{\partial \boldsymbol{\beta}_N} \log f\left(\boldsymbol{\beta}_{0:N} | \boldsymbol{\mathcal{F}}_N\right)\Big|_{\hat{\boldsymbol{\beta}}_{0:N}} \\
&= \left(\boldsymbol{Y}_N - \boldsymbol{z}_N' \hat{\boldsymbol{\beta}}_{N|N}\right)' \boldsymbol{V}_N^{-1} \boldsymbol{z}_N' - \left(\hat{\boldsymbol{\beta}}_{N|N} - \boldsymbol{F}_N \hat{\boldsymbol{\beta}}_{N-1|N}\right)' \boldsymbol{W}_N^{-1} .
\end{aligned}$$

Under the Gaussian assumption (83), the posterior mode and the conditional mean are the same so that the posterior distribution mode estimates $\hat{\boldsymbol{\beta}}_{0:N} = \{\hat{\boldsymbol{\beta}}_{t|N} : t = 0, \ldots, N\}$ are the same as the state space smoothing estimates $\boldsymbol{\beta}_{0:N}^k = \{\boldsymbol{\beta}_{t|N} : t = 0, \ldots, N\}$. The following result provides a direct algebraic proof that the state space smoother estimates maximize the posterior log-likelihood.

Lemma 3.1. *If the Gaussian assumption in (83) is true, then the state space smoothing estimates maximize the posterior log-likelihood*

$$\mathbf{0}_{(N+1 \times p)} = \boldsymbol{\nabla} \ln f\left(\boldsymbol{\beta}_{0:N} | \boldsymbol{\mathcal{F}}_N\right)\big|_{\boldsymbol{\beta}_{0:N}^k} . \qquad (85)$$

Proof: Starting with the Kalman filtering equations, applying the identity $\boldsymbol{z}_N \boldsymbol{V}_N^{-1} = \boldsymbol{P}_{N|N}^{-1} \boldsymbol{K}_N$ and using a little algebra shows

$$\begin{aligned}
\boldsymbol{K}_N \left(\boldsymbol{Y}_N - \boldsymbol{z}_N' \boldsymbol{\beta}_{N|N}\right) &= \left(\boldsymbol{I} - \boldsymbol{K}_N \boldsymbol{z}_N'\right) \left(\boldsymbol{\beta}_{N|N} - \boldsymbol{\beta}_{N|N-1}\right) \\
&= \boldsymbol{P}_{N|N} \boldsymbol{P}_{N|N-1}^{-1} \left(\boldsymbol{\beta}_{N|N} - \boldsymbol{\beta}_{N|N-1}\right) \\
\boldsymbol{z}_N \boldsymbol{V}_N^{-1} \left(\boldsymbol{Y}_N - \boldsymbol{z}_N' \boldsymbol{\beta}_{N|N}\right) &= \boldsymbol{P}_{N|N-1}^{-1} \left(\boldsymbol{\beta}_{N|N} - \boldsymbol{\beta}_{N|N-1}\right) .
\end{aligned}$$

Next, starting with the state space smoothing equations shows

$$\begin{aligned}
\boldsymbol{\beta}_{N|N} - \boldsymbol{F}_N \boldsymbol{\beta}_{N-1|N} &= \left(\boldsymbol{I} - \boldsymbol{F}_N \boldsymbol{\beta}_N\right) \left(\boldsymbol{\beta}_{N|N} - \boldsymbol{\beta}_{N|N-1}\right) \\
&= \boldsymbol{W}_N \boldsymbol{P}_{N|N-1}^{-1} \left(\boldsymbol{\beta}_{N|N} - \boldsymbol{\beta}_{N|N-1}\right) \\
\boldsymbol{W}_N^{-1} \left(\boldsymbol{\beta}_{N|N} - \boldsymbol{F}_N \boldsymbol{\beta}_{N-1|N}\right) &= \boldsymbol{P}_{N|N-1}^{-1} \left(\boldsymbol{\beta}_{N|N} - \boldsymbol{\beta}_{N|N-1}\right) \\
&= \boldsymbol{z}_N \boldsymbol{V}_N^{-1} \left(\boldsymbol{Y}_N - \boldsymbol{z}_N' \boldsymbol{\beta}_{N|N}\right) \ .
\end{aligned}$$

Hence:

$$\mathbf{0} = \left. \frac{\partial}{\partial \boldsymbol{\beta}_N'} \log f\left(\boldsymbol{\beta}_{0:N} \middle| \boldsymbol{\mathcal{F}}_N\right) \right|_{\boldsymbol{\beta}_{0:N}^k} \ . \tag{86}$$

Further analysis of the state space smoothing equations shows

$$\begin{aligned}
\boldsymbol{\beta}_{t|N} - \boldsymbol{F}_t \boldsymbol{\beta}_{t-1|N} &= \left(\boldsymbol{I} - \boldsymbol{F}_t \boldsymbol{B}_t\right) \left(\boldsymbol{\beta}_{t|N} - \boldsymbol{\beta}_{t|t-1}\right) \\
&= \boldsymbol{W}_t \boldsymbol{P}_{t|t-1}^{-1} \left(\boldsymbol{\beta}_{t|N} - \boldsymbol{\beta}_{t|t-1}\right)
\end{aligned}$$

or

$$\begin{aligned}
\boldsymbol{W}_t^{-1} \left(\boldsymbol{\beta}_{t|N} - \boldsymbol{F}_t \boldsymbol{\beta}_{t-1|N}\right) &= \boldsymbol{P}_{t|t-1}^{-1} \left(\boldsymbol{\beta}_{t|N} - \boldsymbol{\beta}_{t|t-1}\right) \\
\boldsymbol{F}_{t+1}' \boldsymbol{W}_{t+1}^{-1} \left(\boldsymbol{\beta}_{t+1|N} - \boldsymbol{F}_{t+1} \boldsymbol{\beta}_{t|N}\right) &= \boldsymbol{F}_{t+1}' \boldsymbol{P}_{t+1|t}^{-1} \left(\boldsymbol{\beta}_{t+1|N} - \boldsymbol{\beta}_{t+1|t}\right) \\
&= \boldsymbol{P}_{t|t}^{-1} \boldsymbol{B}_{t+1} \left(\boldsymbol{\beta}_{t+1|N} - \boldsymbol{\beta}_{t+1|t}\right) \\
&= \boldsymbol{P}_{t|t}^{-1} \left(\boldsymbol{\beta}_{t|N} - \boldsymbol{\beta}_{t|t}\right)
\end{aligned}$$

so that

$$\begin{aligned}
&\boldsymbol{F}_{t+1}' \boldsymbol{W}_{t+1}^{-1} \left(\boldsymbol{\beta}_{t+1|N} - \boldsymbol{F}_{t+1} \boldsymbol{\beta}_{t|N}\right) - \boldsymbol{W}_t^{-1} \left(\boldsymbol{\beta}_{t|N} - \boldsymbol{F}_t \boldsymbol{\beta}_{t-1|N}\right) \\
&= \left(\boldsymbol{P}_{t|t}^{-1} - \boldsymbol{P}_{t|t-1}^{-1}\right) \boldsymbol{\beta}_{t|N} - \boldsymbol{P}_{t|t}^{-1} \boldsymbol{\beta}_{t|t} + \boldsymbol{P}_{t|t-1}^{-1} \boldsymbol{\beta}_{t|t-1} \ .
\end{aligned}$$

Additional analysis of the Kalman Filtering equations shows

$$\mathbf{0} = \boldsymbol{K}_t \left(\boldsymbol{Y}_t - \boldsymbol{z}_t' \boldsymbol{\beta}_{t|N}\right) + \boldsymbol{K}_t \boldsymbol{z}_t' \boldsymbol{\beta}_{t|N} - \boldsymbol{\beta}_{t|t} + \left(\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{z}_t'\right) \boldsymbol{\beta}_{t|t-1}$$

and applying the identities

$$\boldsymbol{P}_{t|t}^{-1} \boldsymbol{K}_t = \boldsymbol{z}_t \boldsymbol{V}_t^{-1} \ \Big| \ \boldsymbol{P}_{t|t}^{-1} \boldsymbol{K}_t \boldsymbol{z}_t' = \boldsymbol{P}_{t|t}^{-1} - \boldsymbol{P}_{t|t-1}^{-1} \ \Big| \ \boldsymbol{P}_{t|t}^{-1} \left(\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{z}_t'\right) = \boldsymbol{P}_{t|t-1}^{-1}$$

shows

$$
\begin{aligned}
\mathbf{0} &= \boldsymbol{z}_t \boldsymbol{V}_t^{-1} \left(\boldsymbol{Y}_t - \boldsymbol{z}_t' \boldsymbol{\beta}_{t|N} \right) \\
&+ \left(\boldsymbol{P}_{t|t}^{-1} - \boldsymbol{P}_{t|t-1}^{-1} \right) \boldsymbol{\beta}_{t|N} - \boldsymbol{P}_{t|t}^{-1} \boldsymbol{\beta}_{t|t} + \boldsymbol{P}_{t|t-1}^{-1} \boldsymbol{\beta}_{t|t-1} \\
&= \boldsymbol{z}_t \boldsymbol{V}_t^{-1} \left(\boldsymbol{Y}_t - \boldsymbol{z}_t' \boldsymbol{\beta}_{t|N} \right) \\
&+ \boldsymbol{F}_{t+1}' \boldsymbol{W}_{t+1}^{-1} \left(\boldsymbol{\beta}_{t+1|t} - \boldsymbol{F}_{t+1} \boldsymbol{\beta}_{t|N} \right) - \boldsymbol{W}_t^{-1} \left(\boldsymbol{\beta}_{t|N} - \boldsymbol{F}_t \boldsymbol{\beta}_{t-1|N} \right) \ .
\end{aligned}
$$

Hence for $t = 1, \ldots, N-1$

$$
\mathbf{0} = \left. \frac{\partial}{\partial \boldsymbol{\beta}_t'} \log f \left(\boldsymbol{\beta}_{0:N} \middle| \boldsymbol{\mathcal{F}}_N \right) \right|_{\boldsymbol{\beta}_{0:N}^k} . \tag{87}
$$

As shown previously by starting with the state space smoothing equations, with initial conditions $\boldsymbol{\beta}_{0|0} = \boldsymbol{b}_0$ and $\boldsymbol{P}_{0|0} = \boldsymbol{W}_0$:

$$
\boldsymbol{F}_1' \boldsymbol{W}_1^{-1} \left(\boldsymbol{\beta}_{1|N} - \boldsymbol{F}_1 \boldsymbol{\beta}_{0|N} \right) = \boldsymbol{P}_{0|0}^{-1} \left(\boldsymbol{\beta}_{0|N} - \boldsymbol{\beta}_{0|0} \right) = \boldsymbol{W}_0^{-1} \left(\boldsymbol{\beta}_{0|N} - \boldsymbol{b}_0 \right)
$$

Hence:

$$
\mathbf{0} = \left. \frac{\partial}{\partial \boldsymbol{\beta}_0'} \log f \left(\boldsymbol{\beta}_{0:N} \middle| \boldsymbol{\mathcal{F}}_N \right) \right|_{\boldsymbol{\beta}_{0:N}^k} . \tag{88}
$$

Intermediate results (86), (87), and (88) prove the desired result (85). ■

The system of state estimating equations associated with the mode of the posterior log-likelihood (84) has the following tridiagonal block matrix representation

$$
\begin{bmatrix}
\boldsymbol{A}_N & -\boldsymbol{C}_N & & & \\
-\boldsymbol{C}_N' & \boldsymbol{B}_{N-1} & -\boldsymbol{C}_{N-1} & & \\
& & \ddots & & \\
& & -\boldsymbol{C}_2' & \boldsymbol{B}_1 & -\boldsymbol{C}_1 \\
& & & -\boldsymbol{C}_1' & \boldsymbol{D}
\end{bmatrix}
\begin{pmatrix}
\boldsymbol{\beta}_{N|N} \\ \boldsymbol{\beta}_{N-1|N} \\ \vdots \\ \boldsymbol{\beta}_{1|N} \\ \boldsymbol{\beta}_{0|N}
\end{pmatrix}
=
\begin{pmatrix}
\boldsymbol{z}_N \boldsymbol{V}_N^{-1} \boldsymbol{Y}_N \\ \boldsymbol{z}_{N-1} \boldsymbol{V}_{N-1}^{-1} \boldsymbol{Y}_{N-1} \\ \vdots \\ \boldsymbol{z}_1 \boldsymbol{V}_1^{-1} \boldsymbol{Y}_1 \\ \boldsymbol{W}_0^{-1} \boldsymbol{b}_0
\end{pmatrix}
$$

where for $t = 1, \ldots, N$

$$
\begin{aligned}
\boldsymbol{A}_N &= \boldsymbol{W}_N^{-1} + \boldsymbol{z}_N \boldsymbol{V}_N^{-1} \boldsymbol{z}_N' \\
\boldsymbol{B}_t &= \boldsymbol{F}_{t+1}' \boldsymbol{W}_{t+1}^{-1} \boldsymbol{F}_{t+1} + \boldsymbol{W}_t^{-1} + \boldsymbol{z}_t \boldsymbol{V}_t^{-1} \boldsymbol{z}_t' \\
\boldsymbol{C}_t &= \boldsymbol{W}_t^{-1} \boldsymbol{F}_t \\
\boldsymbol{D} &= \boldsymbol{F}_1' \boldsymbol{W}_1^{-1} \boldsymbol{F}_1 + \boldsymbol{W}_0^{-1}
\end{aligned}
$$

which is given the following symbolic tridiagonal block representation

$$\boldsymbol{M}_N \boldsymbol{\beta}^k_{N:0} = \boldsymbol{Y}^*_{N:0} \tag{89}$$

where $\boldsymbol{M}_N$ has a tridiagonal block structure with $\mathbf{0}$s in the off tridiagonal block entries. Substituting the actual states $\boldsymbol{\beta}_{N:0} \equiv (\beta_N, \ldots, \beta_0)'$ for the state space smoothers $\boldsymbol{\beta}^k_{N:0}$ in the system of state estimating equations (89) and applying the linear Gaussian state space model (80) shows

$$\boldsymbol{M}_N \boldsymbol{\beta}_{N:0} - \boldsymbol{Y}^*_{N:0} = \begin{pmatrix} \boldsymbol{W}_N^{-1}\boldsymbol{w}_N - \boldsymbol{z}_N \boldsymbol{V}_N^{-1}\boldsymbol{v}_N \\ -\boldsymbol{F}'_N \boldsymbol{W}_N^{-1}\boldsymbol{w}_N + \boldsymbol{W}_{N-1}^{-1}\boldsymbol{w}_{N-1} - \boldsymbol{z}_{N-1}\boldsymbol{V}_{N-1}^{-1}\boldsymbol{v}_{N-1} \\ \vdots \\ -\boldsymbol{F}'_2 \boldsymbol{W}_2^{-1}\boldsymbol{w}_2 + \boldsymbol{W}_1^{-1}\boldsymbol{w}_1 - \boldsymbol{z}_1 \boldsymbol{V}_1^{-1}\boldsymbol{v}_1 \\ -\boldsymbol{F}'_1 \boldsymbol{W}_1^{-1}\boldsymbol{w}_1 + \boldsymbol{W}_0^{-1}\boldsymbol{\beta}_0 \end{pmatrix}$$

which implies the following distribution for the smoother residuals assuming $\boldsymbol{M}_n$ is invertible

$$\begin{aligned} \boldsymbol{M}_N \tilde{\boldsymbol{\beta}}_{N:0|N} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Psi}_N\right) \text{ or } \tilde{\boldsymbol{\beta}}_{N:0|N} &\sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}_N^{-1}\boldsymbol{\Psi}_n \boldsymbol{M}_N^{-1}\right) \\ \tilde{\boldsymbol{\beta}}_{N:0|N} &\equiv \boldsymbol{\beta}_{N:0} - \boldsymbol{\beta}^k_{N:0} \\ &= \left(\beta_t - \beta_{t|N} : t = N, \ldots, 0\right)' \; . \end{aligned}$$

Applying the state space model, where $\{\boldsymbol{\beta}_0\}$, $\{\boldsymbol{w}_t : t = 1, \ldots, N\}$, and $\{\boldsymbol{v}_t : t = 1, \ldots, N\}$ are mutually independent collections of independent random vectors, leads to $\boldsymbol{\Psi}_n = \boldsymbol{M}_n$. Hence

$$\boldsymbol{M}_N \tilde{\boldsymbol{\beta}}_{N:0|N} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}_N\right) \text{ or } \tilde{\boldsymbol{\beta}}_{N:0|N} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}_N^{-1}\right) \; . \tag{90}$$

It is easy to see that $\boldsymbol{\Psi}_N$ also has a tridiagonal block structure if the mutual independence of $\{\boldsymbol{w}_t : t = 1, \ldots, N\}$, and $\{\boldsymbol{v}_t : t = 1, \ldots, N\}$ is relaxed such that $\boldsymbol{w}_{t_1}$ and $\boldsymbol{v}_{t_2}$ are mutually dependent for $t_1 = t_2$ and are mutually independent for $t_1 \neq t_2$ where $t_1, t_2 = 1, \ldots, N$.

One way to solve (89) for the state space smoothers $\boldsymbol{\beta}^k_{N:0}$ is to use the inverse of $\boldsymbol{M}_N$ if the inverse exists

$$\boldsymbol{\beta}^k_{N:0} = \boldsymbol{M}_N^{-1}\boldsymbol{Y}^*_{N:0} \; .$$

Another way to solve (89) for the state space smoothers $\boldsymbol{\beta}^k_{N:0}$ is to use Gaussian elimination to take advantage of the tridiagonal block structure of $\boldsymbol{M}_N$.

Definition 3.1. The likelihood smoother form of the state space smoother is a two pass method for calculating the state space smoother estimates plus a method for calculating the corresponding precision matrices. The first pass consists of using Gaussian elimination to calculate the Kalman filter estimates by removing the upper diagonal of $\boldsymbol{M}_N$ in (89). The second pass consists of using Gaussian elimination to calculate the state space smoother estimates by removing the lower diagonal of $\boldsymbol{M}_N$ in (89). The state space smoother precision matrices are found by using Gaussian elimination to find the diagonal components of $\boldsymbol{M}_N^{-1}$.

Formulas are developed in the next section, for the likelihood smoother estimates and precisions, given a univariate linear Gaussian state space model with constant parameters.

3.4 Asymptotic Precision Analysis

In this section the limiting precision, $\lim_{N\to\infty} P_{t|N}$ for fixed $t \in [1,\ldots,N]$, is investigated for a special case of the Linear Gaussian State Space model:

$$\begin{array}{lll} \text{Initial Information:} & \beta_0 \sim \mathrm{N}(b_0, W_0) & \\ \text{System Equation:} & \beta_t = \phi\beta_{t-1} + w_t, & w_t \sim \mathrm{N}(0, W) \\ \text{Observation Equation:} & Y_t = \eta\beta_t + v_t, & v_t \sim \mathrm{N}(0, V) \end{array} \tag{91}$$

where $\{\beta_0\}$, $\{w_t : t = 1,\ldots,N\}$, and $\{v_t : t = 1,\ldots,N\}$ are mutually independent collections of independent random variables; where the system equation is true for $t = 1,\ldots,N$ and the observation equation is true for all $Y_t \in \boldsymbol{\mathcal{F}}_N$, i.e. for $t = 1,\ldots,N$; where β_t for $t = 0,\ldots,N$ are scalars and $Y_t \in \boldsymbol{\mathcal{F}}_N$ are scalars; and where $|\phi| < 1$ and $|\eta| < 1$.

The system of state estimating equations (89) associated with the above linear Gaussian state space model (91) has the following tridiagonal form

$$\begin{bmatrix} A & -C & & & \\ -C & B & -C & & \\ & & \ddots & & \\ & & -C & B & -C \\ & & & -C & D \end{bmatrix} \begin{pmatrix} \beta_{N|N} \\ \beta_{N-1|N} \\ \vdots \\ \beta_{1|N} \\ \beta_{0|N} \end{pmatrix} = \begin{pmatrix} \frac{\eta}{V} Y_N \\ \frac{\eta}{V} Y_{N-1} \\ \vdots \\ \frac{\eta}{V} Y_1 \\ \frac{b_0}{W_0} \end{pmatrix}$$

where

$$A \equiv \frac{1}{W} + \frac{\eta^2}{V} \qquad\qquad B \equiv \frac{1}{W} + \frac{\phi^2}{W} + \frac{\eta^2}{V}$$

$$C \equiv \frac{\phi}{W} \qquad\qquad D \equiv \frac{1}{W_0} + \frac{\phi^2}{W}$$

which is given the following matrix notation

$$\boldsymbol{M}_N \boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{Y}^*_{N:0|N} \tag{92}$$

where $\boldsymbol{M}_N$ is a tridiagonal matrix with 0s in the off tridiagonal entries, and where $\boldsymbol{\beta}^k_{N:0|N} \equiv (\beta_{N|N}, \ldots, \beta_{0|N})'$ is a vector of the state space smoother estimates for the state vector $\boldsymbol{\beta}_{N:0} \equiv (\beta_N, \ldots, \beta_0)'$ given all of the observations in $\boldsymbol{\mathcal{F}}_N$.

The distribution of the smoother residuals $\tilde{\boldsymbol{\beta}}_{N:0|N} \equiv (\tilde{\beta}_{N|N}, \ldots, \tilde{\beta}_{0|N})'$ from (90) is used to evaluate each precision $P_{t|N} \equiv \mathrm{Var}\ \tilde{\beta}_{t|N}$ for $t = 0, \ldots, N$

$$\boldsymbol{M}_N \tilde{\boldsymbol{\beta}}_{N:0|N} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}_N\right) \text{ or } \tilde{\boldsymbol{\beta}}_{N:0|N} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}_N^{-1}\right) \ .$$

Using the structure of the matrix $\boldsymbol{M}_N$, it is possible to bound each Var $\tilde{\beta}_{t|N}$.

Proposition 3.1. *Given the linear Gaussian state space model defined in (91), then*

$$\begin{aligned}
\left(\frac{1}{W_0} + \frac{\phi^2 + |\phi|}{W}\right)^{-1} &\leq \mathit{Var}\ \tilde{\beta}_{0|N} \leq W_0 \\
\left(\frac{1 + 2|\phi| + \phi^2}{W} + \frac{\eta^2}{V}\right)^{-1} &\leq \mathit{Var}\ \tilde{\beta}_{t|N} \leq \frac{V}{\eta^2},\ t = 1, \ldots, N-1 \\
\left(\frac{1 + |\phi|}{W} + \frac{\eta^2}{V}\right)^{-1} &\leq \mathit{Var}\ \tilde{\beta}_{N|N} \leq \frac{V}{\eta^2}
\end{aligned} \tag{93}$$

.

Proof: The properties of positive definite matrices are used to establish the lower and upper bounds on Var $\tilde{\beta}_{t|N}$ for $t = 1, \ldots, N$. It is easy to show that $\boldsymbol{M}_N$ is positive definite. Let $\boldsymbol{X}_N = (x_N, \ldots, x_0)'$. Then

$$\begin{aligned}
\boldsymbol{X}_N' \boldsymbol{M}_N \boldsymbol{X}_N &= A x_N^2 + \sum_{t=N-1}^{1} B x_t^2 + D x_0^2 - \sum_{t=N}^{1} 2C x_t x_{t-1} \\
&= \sum_{t=N}^{1} \frac{x_t^2 - 2\phi x_t x_{t-1} + \phi^2 x_{t-1}^2}{W} + \frac{\eta^2}{V} x_t^2 + \frac{1}{W_0} x_0^2 \\
&> 0 \text{ for } \boldsymbol{X}_N \neq 0 \ .
\end{aligned}$$

In order to establish the lower bounds in (93) choose $(\rho_N, \rho, \varepsilon)$ as follows

$$\frac{1}{W} - \rho_N \frac{\eta^2}{V} > \frac{|\phi|}{W},\ \frac{1+\phi^2}{W} - \rho \frac{\eta^2}{V} > 2\frac{|\phi|}{W},\ \frac{\phi^2}{W} + \frac{\varepsilon}{W_0} > \frac{|\phi|}{W} \ . \tag{94}$$

and define the following positive definite matrix $M_{(1)}$ as

$$\boldsymbol{M}_{(1)} \equiv \begin{bmatrix} \frac{1}{W} - \rho_N \frac{\eta^2}{V} & C & & & \\ C & \frac{1+\phi^2}{W} - \rho \frac{\eta^2}{V} & C & & \\ & & \ddots & & \\ & & C & \frac{1+\phi^2}{W} - \rho \frac{\eta^2}{V} & C \\ & & & C & \frac{\phi^2}{W} + \frac{\varepsilon}{W_0} \end{bmatrix}.$$

The positive definite property $\boldsymbol{M}_{(2)} \equiv \boldsymbol{M}_N + \boldsymbol{M}_{(1)} > \boldsymbol{M}_N > \mathbf{0}$ implies $\boldsymbol{M}_N^{-1} > \boldsymbol{M}_{(2)}^{-1}$, see Amemiya (1985) [1] Appendix 1 Theorem 17, where

$$\boldsymbol{M}_{(2)} = \begin{bmatrix} 2\frac{1}{W} + (1-\rho_N)\frac{\eta^2}{V} & & & \\ & 2\frac{1+\phi^2}{W} + (1-\rho)\frac{\eta^2}{V} & & \\ & & \ddots & \\ & & & 2\frac{\phi^2}{W} + (1+\varepsilon)\frac{1}{W_0} \end{bmatrix}$$

Note that the positive definite property $\boldsymbol{M}_{(2)} > \boldsymbol{M}_N$ is equivalent to $\boldsymbol{M}_{(2)} - \boldsymbol{M}_N > 0$ where the matrix combination $\boldsymbol{M}_{(2)} - \boldsymbol{M}_N$ is positive definite and where both $\boldsymbol{M}_{(2)}$ and $\boldsymbol{M}_N$ are each positive definite. Hence lower bounds for each Var $\tilde{\beta}_{t|N}$, $t = 0, \ldots, N$, are identified in terms of $(\rho_N, \rho, \varepsilon)$

$$\begin{aligned} \text{Var } \tilde{\beta}_{0|N} &> \left(2\frac{\phi^2}{W} + (1+\varepsilon)\frac{1}{W_0}\right)^{-1} \\ \text{Var } \tilde{\beta}_{t|N} &> \left(2\frac{1+\phi^2}{W} + (1-\rho)\frac{\eta^2}{V}\right)^{-1}, \ t = 1, \ldots, N-1 \\ \text{Var } \tilde{\beta}_{N|N} &> \left(2\frac{1}{W} + (1-\rho_N)\frac{\eta^2}{V}\right)^{-1} . \end{aligned}$$

The desired lower bounds in (93) are found by allowing $(\rho_N, \rho, \varepsilon)$ to change so that the inequalities in (94) converge to equalities.

With regards to the upper bounds in (93), it is convenient to define

$$A(\rho) \equiv \frac{1}{W} + \rho\frac{\eta^2}{V} \tag{95}$$

The analysis proceeds by decomposing $\boldsymbol{X}'_N \boldsymbol{M}_N \boldsymbol{X}_N$ in terms of $A(\rho)$

$$\begin{aligned}
\boldsymbol{X}'_N \boldsymbol{M}_N \boldsymbol{X}_N &= A(\rho) x_N^2 - 2C x_N x_{N-1} + \frac{C^2}{A(\rho)} x_{N-1}^2 + (A - A(\rho))\, x_N^2 \\
&+ \sum_{t=N-1}^{1} A(\rho) x_t^2 - 2C x_t x_{t-1} + \frac{C^2}{A(\rho)} x_{t-1} \\
&+ \sum_{t=N-1}^{1} \left(B - A(\rho) - \frac{C^2}{A(\rho)} \right) x_t^2 \\
&+ \left(D - \frac{C^2}{A(\rho)} \right) x_0^2 \\
&= \boldsymbol{X}'_N \boldsymbol{M}_{(3)} \boldsymbol{X}_N + \boldsymbol{X}'_N \boldsymbol{M}_{(4)} \boldsymbol{X}_N
\end{aligned}$$

where

$$\boldsymbol{M}_{(3)} \equiv \begin{bmatrix} A(\rho) & -C & & & \\ -C & A(\rho) + \frac{C^2}{A(\rho)} & -C & & \\ & & \ddots & & \\ & & -C & A(\rho) + \frac{C^2}{A(\rho)} & -C \\ & & & -C & \frac{C^2}{A(\rho)} \end{bmatrix}$$

$$\boldsymbol{M}_{(4)} \equiv \begin{bmatrix} A - A(\rho) & & & & \\ & B - A(\rho) - \frac{C^2}{A(\rho)} & & & \\ & & \ddots & & \\ & & & B - A(\rho) - \frac{C^2}{A(\rho)} & \\ & & & & D - \frac{C^2}{A(\rho)} \end{bmatrix} .$$

$\boldsymbol{M}_{(3)}$ is positive semi-definite for all values of $\rho \in \mathbb{R}$. $\boldsymbol{M}_{(4)}$ is positive definite for selected values of ρ as follows

$$\begin{aligned}
D - \frac{C^2}{A(\rho)} &= \frac{1}{W_0} + \frac{\phi^2}{W} - \frac{C^2}{A(\rho)} && > 0 \text{ for } \rho \in [0,1) \\
B - A(\rho) - \frac{C^2}{A(\rho)} &= A - A(\rho) + \frac{\phi^2}{W} - \frac{C^2}{A(\rho)} && > 0 \text{ for } \rho \in [0,1) \\
A - A(\rho) &= (1-\rho)\, \frac{\eta^2}{V} && > 0 \text{ for } \rho \in [0,1) \ .
\end{aligned}$$

Consequently $\boldsymbol{M}_N = \boldsymbol{M}_{(3)} + \boldsymbol{M}_{(4)} > \boldsymbol{0}$, $\boldsymbol{M}_N \geq \boldsymbol{M}_{(4)} > \boldsymbol{0}$ implies $\boldsymbol{M}_N^{-1} \leq$

$\boldsymbol{M}_{(4)}^{-1}$. Upper bounds are established in terms of ρ for Var $\tilde{\beta}_{t|N}$, $t = 0, \ldots, N$

$$
\begin{aligned}
\text{Var } \tilde{\beta}_{0|N} &\leq \left(D - \frac{C^2}{A(\rho)} \right)^{-1} \\
\text{Var } \tilde{\beta}_{t|N} &\leq \left(B - A(\rho) - \frac{C^2}{A(\rho)} \right)^{-1}, \; t = 1, \ldots, N-1 \qquad (96) \\
\text{Var } \tilde{\beta}_{N|N} &\leq (A - A(\rho))^{-1} \; .
\end{aligned}
$$

It is easy to show that $(B - A(\rho) - \frac{C^2}{A(\rho)})^{-1}$ and $(A - A(\rho))^{-1}$ are minimized for $\rho \in [0, 1)$ when $\rho = 0$. The upper bounds in (93) are found by choosing $\rho = 0$. ■

It is possible to tighten the upper bounds in (93) by considering two special cases and by continuing to analyze the behavior of the function $A(\rho)$ introduced in Proposition 3.1, see (95).

Proposition 3.2. *Given the linear Gaussian state space model defined in (91)*

$$
\mathit{Var} \; \tilde{\beta}_{0|N} \leq \left(\frac{1}{W_0} + \frac{\phi^2}{W} \frac{\eta^2}{V} \left(\frac{1}{W} + \frac{\eta^2}{V} \right)^{-1} \right)^{-1} \qquad (97)
$$

and if $\phi^2/W + 1/W_0 > |\phi|/W$ *then*

$$
\begin{aligned}
\mathit{Var} \; \tilde{\beta}_{t|N} &\leq \left(\frac{1 - 2|\phi| + \phi^2}{W} + \frac{\eta^2}{V} \right)^{-1}, \; t = 1, \ldots, N-1 \qquad (98) \\
\mathit{Var} \; \tilde{\beta}_{N|N} &\leq \left(\frac{1 - |\phi|}{W} + \frac{\eta^2}{V} \right)^{-1}
\end{aligned}
$$

else if $\phi^2/W + 1/W_0 < |\phi|/W$ *then*

$$
\begin{aligned}
\mathit{Var} \; \tilde{\beta}_{t|N} &\leq \left(\frac{\eta^2}{V} - \frac{1}{W_0} + \frac{1}{W} \frac{1}{W_0} \left(\frac{1}{W_0} + \frac{\phi^2}{W} \right)^{-1} \right)^{-1}, \; t = 1, \ldots, N-1 \\
& \qquad (99) \\
\mathit{Var} \; \tilde{\beta}_{N|N} &\leq \left(\frac{\eta^2}{V} + \frac{1}{W} \frac{1}{W_0} \left(\frac{1}{W_0} + \frac{\phi^2}{W} \right)^{-1} \right)^{-1} \; .
\end{aligned}
$$

Proof: With regards to the upper bounds in (98), assume that $\phi^2/W + 1/W_0 > |\phi|/W$, and choose $(\rho_N, \rho, \varepsilon)$ so that the inequalities in (94) are satisfied and $1-\varepsilon>0$. Define the following positive definite matrices

$$\boldsymbol{M}_{(5)} \equiv \begin{bmatrix} \frac{1}{W}-\rho_N\frac{\eta^2}{V} & -C & & & \\ -C & \frac{1+\phi^2}{W}-\rho\frac{\eta^2}{V} & -C & & \\ & & \ddots & & \\ & & -C & \frac{1+\phi^2}{W}-\rho\frac{\eta^2}{V} & -C \\ & & & -C & \frac{\phi^2}{W}+\frac{\varepsilon}{W_0} \end{bmatrix}$$

$$\boldsymbol{M}_{(6)} \equiv \begin{bmatrix} (1+\rho_N)\,\frac{\eta^2}{V} & & & & \\ & (1+\rho)\,\frac{\eta^2}{V} & & & \\ & & \ddots & & \\ & & & (1+\rho)\,\frac{\eta^2}{V} & \\ & & & & \frac{1-\varepsilon}{W_0} \end{bmatrix}$$

such that $\boldsymbol{M}_N=\boldsymbol{M}_{(5)}+\boldsymbol{M}_{(6)}>\boldsymbol{0}$, $\boldsymbol{M}_N>\boldsymbol{M}_{(6)}>\boldsymbol{0}$, and $\boldsymbol{M}_N^{-1}<\boldsymbol{M}_{(6)}^{-1}$. Hence upper bounds for Var $\tilde{\beta}_{t|N}$ are identified in terms of $(\rho_N,\rho,\varepsilon)$ for $t=0,\ldots,N$

$$\begin{aligned} \text{Var }\tilde{\beta}_{0|N} &< \frac{W_0}{1-\varepsilon} \\ \text{Var }\tilde{\beta}_{t|N} &< (1+\rho)^{-1}\,\frac{V}{\eta^2},\ t=1,\ldots,N-1 \\ \text{Var }\tilde{\beta}_{N|N} &< (1+\rho_N)^{-1}\,\frac{V}{\eta^2}\,. \end{aligned}$$

The desired upper bounds in (98) for each Var $\tilde{\beta}_{t|N}$, $t=1,\ldots,N$, are found by allowing $(\rho_N,\rho,\varepsilon)$ to change such that the inequalities in (94) converge to equalities. The corresponding upper bound for Var $\tilde{\beta}_{0|N}$ is $(1/W_0+(\phi^2-|\phi|)/W)^{-1}$.

With regard to the upper bounds in (99), the upper bounds in (96) as a function of $A(\rho)$ are analyzed for $\rho\in(-\infty,1)$. The function $A(\rho)$ has a local maximum, a local minimum, and a singularity point between the local maximum and the local minimum. Let ρ_1 denote the local maximum, let ρ_2 denote the local minimum, and let ρ_s denote the singularity point

$$\rho_1=-\left(1+|\phi|\right)\frac{V}{\eta^2}\frac{1}{W}<\rho_s=-\frac{V}{\eta^2}\frac{1}{W}<\rho_2=-\left(1-|\phi|\right)\frac{V}{\eta^2}\frac{1}{W}\,.$$

If $(D-C^2/A(\rho_2))>0$ then the upper bounds in (98) are valid; otherwise, different upper bounds are found by decreasing ρ from 0 such that $(D-$

$C^2/A(\rho)) \to 0$. Let ρ_3 denote the value of ρ such that $(D - C^2/A(\rho_3)) = 0$

$$\rho_s < \rho_3 = -\frac{1}{W_0}\frac{1}{W}\frac{V}{\eta^2}\left(\frac{1}{W_0} + \frac{\phi^2}{W}\right)^{-1} .$$

It is easy to show that $\rho_3 < \rho_2$ is equivalent to $\phi^2/W + 1/W_0 > |\phi|/W$. For $\rho_2 < \rho_3$, the upper bounds in (99) for each Var $\tilde{\beta}_{t|N}$, $t = 1, \ldots, N$, are found by allowing $\rho \to \rho_3$. As $\rho \to \rho_3$ the corresponding upper bound on Var $\tilde{\beta}_{0|N}$ is $+\infty$.

The bound $(D - C^2/A(\rho))^{-1}$ on Var $\tilde{\beta}_{0|N}$ for $\rho \in (-\infty, 1)$ is minimized as ρ approaches 1. The upper bound in (97) for Var $\tilde{\beta}_{0|N}$ is found by allowing $\rho \to 1$. ■

Note that the bounds for each Var $\tilde{\beta}_{t|N}$, $t \in [0, \ldots, N]$ identified in Proposition 3.2 can be shown by direct computation to be tighter or equal to the bounds provided in Proposition 3.1.

Corollary 3.1. *If $\{\beta_t : t = 0, \ldots, N\}$ is stationary, i.e. $b_0 = 0$ and $W_0 = W/(1-\phi^2)$, then direct calculation shows that $\phi^2/W + 1/W_0 = 1/W > |\phi|/W$ for $|\phi| < 1$. Hence the first set of bounds (98) in Proposition 3.2 apply.* ■

Next, the tri-diagonal property of the $\boldsymbol{M}_N$ matrix is exploited to provide simple formulas for the state space smoothers in $\boldsymbol{\beta}^k_{N:0|N}$ and for the elements of $\boldsymbol{M}_N^{-1}$ that correspond to the covariances of $(\tilde{\beta}_{t_1|N}, \tilde{\beta}_{t_2|N})$, $t_1, t_2 = 0, \ldots, N$.

Proposition 3.3. *Given the linear Gaussian state space model defined in (91) then the Kalman filter and smoother estimates are calculated as follows*

$$\begin{aligned}
\beta_{0|0} &\equiv b_0 \\
\beta_{1|1} &= \left(A - \frac{C^2}{G_1^*}\right)^{-1}\left(\frac{\eta}{V}Y_1 + \frac{C}{G_1^*}\frac{1}{W_0}\beta_{0|0}\right) \\
\beta_{t|t} &= \left(A - \frac{C^2}{G_t^*}\right)^{-1}\left(\frac{\eta}{V}Y_t + \frac{C}{G_t^*}\left(A - \frac{C^2}{G_{t-1}^*}\right)\beta_{t-1|t-1}\right),\ t = 2, \ldots, N \\
\beta_{t|N} &= \frac{1}{G_{t+1}^*}\left(A - \frac{C^2}{G_t^*}\right)\beta_{t|t} + \frac{C}{G_{t+1}^*}\beta_{t+1|N},\ t = N-1, \ldots, 1 \\
\beta_{0|N} &= \frac{1}{G_1^*}\frac{1}{W_0}\beta_{0|0} + \frac{C}{G_1^*}\beta_{1|N}
\end{aligned}$$

where

$$G_j^* \equiv \begin{cases} D & : j = 1 \\ B - \frac{C^2}{G_{j-1}^*} & : j > 1 \end{cases} .$$

Proof: Gaussian elimination of $\boldsymbol{M}_N\boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{Y}^*_{N:0|N}$ for $N = 1, 2$ to remove the upper diagonal in $\boldsymbol{M}_N$ shows that the Kalman filter estimates are

$$
\begin{aligned}
\beta_{0|0} &\equiv b_0 \\
\beta_{1|1} &= \left(A - \frac{C^2}{G_1^*}\right)^{-1} \left(\frac{\eta}{V}Y_1 + \frac{C}{G_1^*}\frac{b_0}{W_0}\right) \\
\beta_{2|2} &= \left(A - \frac{C^2}{G_2^*}\right)^{-1} \left(\frac{\eta}{V}Y_2 + \frac{C}{G_2^*}\left(\frac{\eta}{V}Y_1 + \frac{C}{G_1^*}\frac{b_0}{W_0}\right)\right) .
\end{aligned}
$$

Induction shows the following formulas for the Kalman filter estimates at time indices $t-1$ and t for $t > 1$

$$
\begin{aligned}
\beta_{t-1|t-1} &= \left(A - \frac{C^2}{G_{t-1}^*}\right)^{-1} \left(\frac{\eta}{V}Y_{t-1} + \frac{C}{G_{t-1}^*}\left(\frac{\eta}{V}Y_{t-2} + \dots \right.\right. \\
&\quad \left.\left. + \frac{C}{G_2^*}\left(\frac{\eta}{V}Y_1 + \frac{C}{G_1^*}\frac{b_0}{W_0}\right)\dots\right)\right) \\
\beta_{t|t} &= \left(A - \frac{C^2}{G_t^*}\right)^{-1} \left(\frac{\eta}{V}Y_t + \frac{C}{G_t^*}\left(\frac{\eta}{V}Y_{t-1} + \dots \right.\right. \\
&\quad \left.\left. + \frac{C}{G_2^*}\left(\frac{\eta}{V}Y_1 + \frac{C}{G_1^*}\frac{b_0}{W_0}\right)\dots\right)\right) .
\end{aligned}
$$

The previous display is used to prove the recursive formula result for the Kalman filter estimate $\beta_{t|t}$ given $\beta_{t-1|t-1}$ with $t > 1$.

Gaussian elimination of $\boldsymbol{M}_N\boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{Y}^*_{N:0|N}$ for $N = 1, 2$ to remove the upper diagonal in $\boldsymbol{M}_N$ results in the following system of equations

$$
\begin{bmatrix}
1 & & & & \\
-\frac{C}{G_N^*} & 1 & & & \\
 & & \ddots & & \\
 & & -\frac{C}{G_2^*} & 1 & \\
 & & & -\frac{C}{G_1^*} & 1
\end{bmatrix}
\begin{pmatrix}
\beta_{N|N} \\ \beta_{N-1|N} \\ \vdots \\ \beta_{1|N} \\ \beta_{0|N}
\end{pmatrix}
=
\begin{pmatrix}
\beta_{N|N} \\
\frac{1}{G_N^*}\left(A - \frac{C}{G_{N-1}^*}\right)\beta_{N-1|N-1} \\
\vdots \\
\frac{1}{G_2^*}\left(A - \frac{C}{G_1^*}\right)\beta_{1|1} \\
\frac{1}{G_1^*}\frac{1}{W_0}\beta_{0|0}
\end{pmatrix}
$$

The previous display is used to prove the recursive formula result for the state space smoother estimate $\beta_{t|N}$ given $\beta_{t+1|N}$ and given the Kalman filter estimate $\beta_{t|t}$ with $t > 1$. Hence the complete result is proven. ■

Lemma 3.2. *Given the linear Gaussian state space model defined in (91)*

$$\begin{aligned}
Var\ \tilde{\beta}_{0|N} &= \left(D - \frac{C^2}{G_N}\right)^{-1} \\
Cov\left(\tilde{\beta}_{0|N}, \tilde{\beta}_{t|N}\right) &= \frac{C}{G_{N-t+1}} \times \cdots \times \frac{C}{G_N}\, Var\ \tilde{\beta}_{0|N},\ t = 1, \ldots, N \\
G_j &\equiv \begin{cases} A & : j = 1 \\ B - \frac{C^2}{G_{j-1}} & : j > 1 \end{cases} .
\end{aligned}$$

Proof: Gaussian elimination is used to solve $\boldsymbol{M}_N \boldsymbol{X}_N = \boldsymbol{e}_{N+1}$ where $\boldsymbol{e}_{N+1} = (0, \ldots, 0, 1)'$. The Gaussian elimination of $\boldsymbol{M}_N$ proceeds by eliminating the lower diagonal starting from the left and then by eliminating the upper diagonal starting from the right. For $N = 3$ the resulting solution for $\boldsymbol{X}_3$ is

$$\boldsymbol{X}_3 = \begin{pmatrix} x_3 \\ x_2 \\ x_1 \\ x_0 \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}\left(\tilde{\beta}_{0|N}, \tilde{\beta}_{3|N}\right) \\ \mathrm{Cov}\left(\tilde{\beta}_{0|N}, \tilde{\beta}_{2|N}\right) \\ \mathrm{Cov}\left(\tilde{\beta}_{0|N}, \tilde{\beta}_{1|N}\right) \\ \mathrm{Var}\ \tilde{\beta}_{0|N} \end{pmatrix} = \begin{pmatrix} \frac{C}{G_1} x_2 \\ \frac{C}{G_2} x_1 \\ \frac{C}{G_3} x_0 \\ \left(D - \frac{C^2}{G_3}\right)^{-1} \end{pmatrix} .$$

Generalizing the result in the previous display for $N > 3$ proves the result. ■

Corollary 3.2. *If $\{\beta_t : t = 0, \ldots, N\}$ is stationary, i.e. $b_0 = 0$ and $W_0 = W/(1 - \phi^2)$, then direct calculation shows $D = 1/W$ and $G_2^* = A$. Hence $G_j^* = G_{j-1}$ for $j \in \{2, 3, \ldots\}$.* ■

Lemma 3.3. *Given the linear Gaussian state space model defined in (91)*

$$\begin{aligned}
Var\ \tilde{\beta}_{N|N} &= \left(A - \frac{C^2}{G_N^*}\right)^{-1} \\
Cov\left(\tilde{\beta}_{t|N}, \tilde{\beta}_{N|N}\right) &= \frac{C}{G_{t+1}^*} \times \cdots \times \frac{C}{G_N^*}\, Var\ \tilde{\beta}_{N|N},\ t = 0, \ldots, N-1 .
\end{aligned}$$

Proof: Gaussian elimination is used to solve $\boldsymbol{M}_N \boldsymbol{X}_N = \boldsymbol{e}_1$ where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$. The Gaussian elimination of $\boldsymbol{M}_N$ proceeds by eliminating the upper diagonal starting from the right and then by eliminating the lower diagonal starting from the left. For $N = 3$ the resulting solution for $\boldsymbol{X}_N$ is

$$\boldsymbol{X}_3 = \begin{pmatrix} x_3 \\ x_2 \\ x_1 \\ x_0 \end{pmatrix} = \begin{pmatrix} \text{Var } \tilde{\beta}_{3|N} \\ \text{Cov}\left(\tilde{\beta}_{2|N}, \tilde{\beta}_{3|N}\right) \\ \text{Cov}\left(\tilde{\beta}_{1|N}, \tilde{\beta}_{3|N}\right) \\ \text{Cov}\left(\tilde{\beta}_{0|N}, \tilde{\beta}_{3|N}\right) \end{pmatrix} = \begin{pmatrix} \left(A - \frac{C^2}{G_3^*}\right)^{-1} \\ \frac{C}{G_3^*} x_3 \\ \frac{C}{G_2^*} x_2 \\ \frac{C}{G_1^*} x_1 \end{pmatrix} .$$

Generalizing the result in the previous display for $N > 3$ proves the result. ■

Lemma 3.4. *Given the linear Gaussian state space model defined in (91)*

$$\begin{aligned} Var\ \tilde{\beta}_{t|N} &= \left(G_{N-t+1} - \frac{C^2}{G_t^*}\right)^{-1},\ t = 1, \ldots, N-1 \\ Cov\left(\tilde{\beta}_{t_1|N}, \tilde{\beta}_{t|N}\right) &= \frac{C}{G_{t_1+1}^*} \times \cdots \times \frac{C}{G_t^*} Var\ \tilde{\beta}_{t|N},\ t_1 = 0, \ldots, t-1 \\ Cov\left(\tilde{\beta}_{t|N}, \tilde{\beta}_{t_1|N}\right) &= \frac{C}{G_{N-t_1+1}} \times \cdots \times \frac{C}{G_{N-t}} Var\ \tilde{\beta}_{t|N},\ t_1 = t+1, \ldots, N . \end{aligned}$$

Proof: Given a fixed $t \in [1, \ldots, N-1]$, Gaussian elimination is used to solve $\boldsymbol{M}_N \boldsymbol{X}_N = \boldsymbol{e}_{N-t+1}$ where $\boldsymbol{e}_{N-t+1}$ is a vector consisting of $N+1$ zeros except for a one in element number $N-t+1$. The Gaussian elimination of $\boldsymbol{M}_N$ proceeds by eliminating $N-t$ elements in the lower diagonal starting from the left and then by eliminating t elements in the upper diagonal starting from the right. The remainder of the elements in the upper and lower diagonals are then eliminated. For $N = 3$ and $t = 2$ the resulting solution for $\boldsymbol{X}_3$ is

$$\boldsymbol{X}_3 = \begin{pmatrix} x_3 \\ x_2 \\ x_1 \\ x_0 \end{pmatrix} = \begin{pmatrix} \text{Cov}\left(\tilde{\beta}_{2|N}, \tilde{\beta}_{3|N}\right) \\ \text{Var } \tilde{\beta}_{2|N} \\ \text{Cov}\left(\tilde{\beta}_{1|N}, \tilde{\beta}_{2|N}\right) \\ \text{Cov}\left(\tilde{\beta}_{0|N}, \tilde{\beta}_{2|N}\right) \end{pmatrix} = \begin{pmatrix} \frac{C}{G_1} x_2 \\ \left(B - \frac{C^2}{G_1} - \frac{C^2}{G_2^*}\right)^{-1} \\ \frac{C}{G_2^*} x_2 \\ \frac{C}{G_1^*} x_1 \end{pmatrix} .$$

Generalizing the result in the previous display for $N > 3$ proves the result. ■

Corollary 3.3. *Given the linear Gaussian state space model defined in (91)*

$$Var\ \tilde{\beta}_{t|N} = \frac{G_t^*}{G_{N-t+1}} \times \cdots \times \frac{G_1^*}{G_N} Var\ \tilde{\beta}_{0|N},\ t = 1, \ldots, N .$$

Proof: The following variance ratio equation is shown by simple algebra for $t = 2, \dots, N-1$

$$\frac{\text{Var }\tilde{\beta}_{t|N}}{\text{Var }\tilde{\beta}_{t-1|N}} = \frac{G_t^* - \frac{C^2}{G_{N-t+1}}}{G_{N-t+1} - \frac{C^2}{G_t^*}} = \frac{G_t^*}{G_{N-t+1}} \; .$$

Generalizing the previous display for $t = 1$ and $t = N$ proves the result. ■

Remark 3.1. The vector of state space smoother estimates can be calculated using $\boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{M}_N^{-1}\boldsymbol{Y}^*_{N:0|N}$ since $\boldsymbol{M}_N$ is positive definite as shown in Proposition 3.1 and is invertible. As shown in Proposition 3.3 and Lemmas 3.2 through 3.4, Gaussian elimination can be used to solve $\boldsymbol{M}_N\boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{Y}^*_{N:0|N}$ for $\boldsymbol{\beta}^k_{N:0|N}$ and to invert $\boldsymbol{M}_N$ for the smoother precisions in Var $\boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{M}_N^{-1}$. The likelihood smoother form of the state space smoother consists of a two pass method to calculate the state space smoother estimates and a method to calculate the state space smoother precisions. The first pass of the likelihood smoother estimate method calculates

$$\begin{aligned}
G_j^* &\equiv \begin{cases} D & : j = 1 \\ B - \frac{C^2}{G_{j-1}^*} & : j > 1 \end{cases} \quad \text{for } j = 1, \dots, N \\
\beta^*_{0|0} &= \frac{1}{W_0}\beta_{0|0} = \frac{1}{W_0}b_0 \\
\beta^*_{t|t} &= \left(A - \frac{C^2}{G_t^*}\right)\beta_{t|t} = \frac{\eta}{V}Y_t + \frac{C}{G_t^*}\beta^*_{t-1|t-1} \text{ for } t = 1, \dots, N \; .
\end{aligned}$$

The second pass of the likelihood smoother estimate method calculates

$$\begin{aligned}
\beta_{N|N} &= \left(A - \frac{C^2}{G_N^*}\right)^{-1}\beta^*_{N|N} \\
\beta_{t|N} &= \frac{1}{G_{t+1}^*}\left(\beta^*_{t|t} + C\beta_{t+1|N}\right) \text{ for } t = N-1, \dots, 0 \; .
\end{aligned}$$

The likelihood smoother precision method calculates

$$G_j \equiv \begin{cases} A & : j = 1 \\ B - \frac{C^2}{G_{j-1}} & : j > 1 \end{cases} . \text{ for } j = 1, \dots, N$$

$$P_{N|N} = \text{Var } \tilde{\beta}_{N|N} = \left(A - \frac{C^2}{G_N^*} \right)^{-1}$$

$$P_{t|N} = \text{Var } \tilde{\beta}_{t|N} = \left(G_{N-t+1} - \frac{C^2}{G_t^*} \right)^{-1} \quad \text{for } t = N-1, \dots, 1$$

$$P_{0|N} = \text{Var } \tilde{\beta}_{0|N} = \left(D - \frac{C^2}{G_N} \right)^{-1} .$$

The first pass is equivalent to performing Kalman prediction and filtering to obtain $\beta_{N|N}$ and the second pass calculates the state space smoother estimates $\beta_{t|N}$ for $t = N, \dots, 0$ based on the first pass. When new observations become available, then only the end of the first pass and the complete second pass of the likelihood smoother estimate method as well as the likelihood smoother precision method need to be redone. Note that an alternative Gaussian elimination procedure can be used to solve $\boldsymbol{M}_N \boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{Y}^*_{N:0}$ for $\boldsymbol{\beta}^k_{N:0|N}$ by first removing the lower diagonal of $\boldsymbol{M}_N$ and then removing the upper diagonal of $\boldsymbol{M}_N$. This alternative Gaussian elimination procedure is less efficient than the likelihood smoother estimate method introduced above in the sense that the alternative Gaussian elimination procedure would have to be redone in total when new observations become available.

Before establishing the limit as $N \to \infty$ for Var $\tilde{\beta}_{t|N}$, $t \in [0, \dots, N]$, the behavior of G_j and G_j^* is established as $l \to \infty$

Lemma 3.5. *The properties of G_j defined in Lemma 3.2 include*

$$G_j \to G_\infty = \frac{B + \sqrt{B^2 - 4C^2}}{2} \text{ as } j \to \infty \tag{100}$$

$$A \le G_j < G_{j+1} < G_\infty, \ j = 1, 2, \dots . \tag{101}$$

Proof: The following bounds is used to prove (101)

$$A \le G_j < G_{j+1} < B, \ j = 1, 2, \dots . \tag{102}$$

By direct calculation $A + C^2/A < B$ proves (102) for $j = 1$. The general result (102) for $j > 1$ is proven by induction. Hence $G_j \to G_\infty$ as $j \to \infty$.

At convergence G_∞ has two possible solutions

$$G_\infty = B - \frac{C^2}{G_\infty} \Leftrightarrow G_\infty^2 - BG_\infty + C^2 = 0$$
$$G_\infty = \frac{B \pm \sqrt{B^2 - 4C^2}}{2} .$$

Direct calculation shows that the larger solution for G_∞ identified in (100) is the only solution that satisfies (102) such that $A < G_\infty$. Induction is used to prove (101). ■

Lemma 3.6. *The properties of G_j^* defined in Proposition 3.3 include*

$$G_j^* \to G_\infty^* = \frac{B + \sqrt{B^2 - 4C^2}}{2} = G_\infty \text{ as } j \to \infty \tag{103}$$
$$\text{If } D < G_\infty^* \text{ then } D \le G_j^* < G_{j+1}^* < G_\infty^*,\ j = 1, 2, \dots \tag{104}$$
$$\text{If } G_\infty^* < D \text{ then } G_\infty^* < G_{j+1}^* < G_j^* \le D,\ j = 1, 2, \dots . \tag{105}$$

Proof: By direct calculation $C^2/A < D$ and $C^2/A < G_2^* < B$. Induction for $j > 2$ is used to show the general result that

$$\frac{C^2}{A} < G_j^* < B,\ j = 2, 3, \dots \tag{106}$$

If $G_1^* < G_2^*$ then induction shows $C^2/A < G_j^* < G_{j+1}^* < B$ for $j = 1, 2, \dots$. If $G_2^* < G_1^*$ then induction shows $C^2/A < G_{j+1}^* < G_j^* < B$ for $j = 2, 3, \dots$. Hence $G_j^* \to G_\infty^*$ as $j \to \infty$. At convergence G_∞^* has two possible solutions

$$G_\infty^* = B - \frac{C^2}{G_\infty^*} \Leftrightarrow (G_\infty^*)^2 - BG_\infty^* + C^2 = 0$$
$$G_\infty^* = \frac{B \pm \sqrt{B^2 - 4C^2}}{2} .$$

Direct calculation shows that the larger solution for G_∞^* identified in (103) is the only solution that satisfies (106) such that $C^2/A < G_\infty^*$. Induction is used to prove (104) and (105). ■

Limits and bounds on each Var $\tilde{\beta}_{t|N}$, $t \in [0, \dots, N]$ are now established using Lemmas 3.2 through 3.6.

Theorem 3.1. *Limits for each Var* $\tilde{\beta}_{t|N}$, $t \in [0, \ldots, N]$ *as* $N \to \infty$ *are*

$$\begin{aligned} Var\ \tilde{\beta}_{0|N} &\to \left(D - \frac{C^2}{G_\infty}\right)^{-1} \\ Var\ \tilde{\beta}_{t|N} &\to \left(G_\infty - \frac{C^2}{G_t^*}\right)^{-1}, \ \textit{for fixed } t \in [1, \ldots, \infty) \\ Var\ \tilde{\beta}_{N|N} &\to \left(A - \frac{C^2}{G_\infty^*}\right)^{-1} . \end{aligned}$$

Var $\tilde{\beta}_{0|N}$ *is bounded as follows*

$$\left(D - \frac{C^2}{G_\infty}\right)^{-1} < Var\ \tilde{\beta}_{0|N} < \left(D - \frac{C^2}{A}\right)^{-1} .$$

If $D < G_\infty$ *then bounds on each Var* $\tilde{\beta}_{t|N}$, $t \in [1, \ldots, N]$ *are as follows*

$$\begin{aligned} \left(G_\infty - \frac{C^2}{G_\infty^*}\right)^{-1} &< Var\ \tilde{\beta}_{t|N} < \left(G_2 - \frac{C^2}{D}\right)^{-1}, \ t \in [1, \ldots, N-1] \\ \left(A - \frac{C^2}{G_\infty^*}\right)^{-1} &< Var\ \tilde{\beta}_{N|N} < \left(A - \frac{C^2}{D}\right)^{-1} \end{aligned}$$

else if $D > G_\infty$ *then bounds on each Var* $\tilde{\beta}_{t|N}$, $t \in [1, \ldots, N]$ *are as follows*

$$\begin{aligned} \left(G_\infty - \frac{C^2}{D}\right)^{-1} &< Var\ \tilde{\beta}_{t|N} < \left(G_2 - \frac{C^2}{G_\infty^*}\right)^{-1}, \ t \in [1, \ldots, N-1] \\ \left(A - \frac{C^2}{D}\right)^{-1} &< Var\ \tilde{\beta}_{N|N} < \left(A - \frac{C^2}{G_\infty^*}\right)^{-1} . \blacksquare \end{aligned}$$

Note that the bounds for each Var $\tilde{\beta}_{t|N}$, $t \in [0, \ldots, N]$ identified in Theorem 3.1 can be shown by direct computation to be tighter or equal to the bounds provided in Propositions 3.1 and 3.2.

The following corollary provides the asymptotic precision for $P_{t|N}$ where t no longer remains fixed as a function of N, for example $t \equiv t(N) = kN$ where $k \in (0, 1)$.

Corollary 3.4. *If* $t \equiv t(N)$ *such that* $t(N) \to \infty$ *and* $N - t(N) \to \infty$ *as* $N \to \infty$ *then* $P_{t(N)|N} \to (G_\infty - C^2/G_\infty^*)^{-1} = (2G_\infty - B)^{-1}$ *as* $N \to \infty$.

As a check on the precision $P_{N|N} =$ Var $\tilde{\beta}_{N|N}$, the following corollary shows that the equation for $P_{N|N}$ satisfies the Kalman predictor and filter methods.

Corollary 3.5. *The equation for $P_{N|N} = Var\ \tilde{\beta}_{N|N}$ from Lemma 3.3 satisfies the Kalman predictor and filter methods such that*

$$P_{N|N} = \frac{VP_{N|N-1}}{\eta^2 P_{N|N-1} + V},\ P_{N|N-1} = \phi^2 P_{N-1|N-1} + W\ .$$

Proof: Inverting the equation for $P_{N|N}$ in the previous display with respect to $P_{N-1|N-1}$ shows

$$\begin{aligned} P_{N-1|N-1} &= \frac{1}{\phi^2}\left(\frac{VP_{N|N}}{V - \eta^2 P_{N|N}} - W\right) \\ &= \frac{W}{\phi^2}\left(\frac{\frac{1}{W}}{P_{N|N}^{-1} - \frac{\eta^2}{V}} - 1\right)\ . \end{aligned}$$

Inserting the equations for $P_{N-1|N-1} = \text{Var}\ \tilde{\beta}_{N-1|N-1}$ and $P_{N|N} = \text{Var}\ \tilde{\beta}_{N|N}$ from Lemma 3.3 into the left and right hand sides of the previous display and reducing shows

$$\begin{aligned} \text{l.h.s.} &= \left(A - \frac{C^2}{G^*_{N-1}}\right)^{-1} \\ \text{r.h.s.} &= \left(G^*_N - \frac{\phi^2}{W}\right)^{-1}\ . \end{aligned}$$

Hence the result is proven since $G^*_N = B - C^2/G^*_{N-1}$ from Proposition 3.3. ■

The following proposition shows how the asymptotic filter precision satisfies the steady state Riccati equation, see [29] section 4.3.

Proposition 3.4. *The asymptotic one step ahead predictor precision $P_{+1} = \phi^2 P + W$ satisfies the steady state Riccati equation where $P = \lim_{N\to\infty} P_{N|N} = (G_\infty - \phi^2/W)^{-1}$ identifies the asymptotic filter precision*

$$P_{+1} = \phi^2\left(1 - \eta^2 P_{+1}\left(\eta^2 P_{+1} + V\right)^{-1}\right) P_{+1} + W\ .$$

Proof: Algebraic manipulation of the steady state Riccati equation shows that P_{+1} is a zero of the following quadratic equation

$$\frac{1}{W}\frac{\eta^2}{V}P_{+1}^2 + \left(\frac{1}{W} - \frac{\phi^2}{W} - \frac{\eta^2}{V}\right)P_{+1} - 1 = 0\ .$$

Hence P_{+1} has two possible roots

$$P_{+1} = \frac{-\left(\frac{1}{W} - \frac{\phi^2}{W} - \frac{\eta^2}{V}\right) \pm \sqrt{\left(\frac{1}{W} - \frac{\phi^2}{W} - \frac{\eta^2}{V}\right)^2 + 4\frac{1}{W}\frac{\eta^2}{V}}}{2\frac{1}{W}\frac{\eta^2}{V}} \, .$$

It will be shown that the larger of the two possible roots is the correct value.

The asymptotic filter precision P satisfies

$$\begin{aligned} P^{-1} &= A - \frac{C^2}{G_\infty} = G_\infty - \frac{\phi^2}{W} \\ &= \frac{\left(\frac{1}{W} - \frac{\phi^2}{W} + \frac{\eta^2}{V}\right) + \sqrt{\left(\frac{1}{W} - \frac{\phi^2}{W} + \frac{\eta^2}{V}\right)^2 + 4\frac{\phi^2}{W}\frac{\eta^2}{V}}}{2} \, . \end{aligned}$$

Hence P^{-1} is a zero of the following quadratic equation

$$P^{-2} - \left(\frac{1}{W} - \frac{\phi^2}{W} + \frac{\eta^2}{V}\right) P^{-1} - \frac{\phi^2}{W}\frac{\eta^2}{V} = 0 \, .$$

The above quadratic equation can also be derived by starting with the steady state equation for the asymptotic filter precision, see [7] section 4.2.3.

$$P = \frac{\phi^2 P + W}{\frac{\eta^2}{V}\left(\phi^2 P + W\right) + 1}$$

and deriving the following quadratic equation in P

$$\frac{\phi^2}{W}\frac{\eta^2}{V} P^2 + \left(\frac{1}{W} - \frac{\phi^2}{W} + \frac{\eta^2}{V}\right) P - 1 = 0 \, .$$

The larger of the two possible roots of the previous quadratic equation in P satisfies $P \times P^{-1} = 1$ where $P^{-1} = G_\infty - \phi^2/W$ from above and where

$$\begin{aligned} P &= \frac{-\left(\frac{1}{W} - \frac{\phi^2}{W} + \frac{\eta^2}{V}\right) + \sqrt{\left(\frac{1}{W} - \frac{\phi^2}{W} + \frac{\eta^2}{V}\right)^2 + 4\frac{\phi^2}{W}\frac{\eta^2}{V}}}{2\frac{\phi^2}{W}\frac{\eta^2}{V}} \\ &= -\left(\frac{\phi^2}{W}\frac{\eta^2}{V}\right)^{-1} \left(\frac{B - \sqrt{B^2 - 4C^2}}{2} - \frac{\phi^2}{W}\right) \, . \end{aligned}$$

Hence the asymptotic one step ahead predictor precision satisfies

$$\begin{aligned} P_{+1} &= \phi^2 P + W \\ &= \frac{-\left(\frac{1}{W} - \frac{\phi^2}{W} - \frac{\eta^2}{V}\right) + \sqrt{\left(\frac{1}{W} - \frac{\phi^2}{W} - \frac{\eta^2}{V}\right)^2 + 4\frac{1}{W}\frac{\eta^2}{V}}}{2\frac{1}{W}\frac{\eta^2}{V}} \, . \blacksquare \end{aligned}$$

Remark 3.2. The results of this section can be generalized to the following linear state space model where the Gaussian assumption has been removed

$$
\begin{array}{lll}
\text{Initial Information:} & \beta_0 \sim (b_0, W_0) & \\
\text{System Equation:} & \beta_t = \phi\beta_{t-1} + w_t, & w_t \sim (0, W) \\
\text{Observation Equation:} & Y_t = \eta\beta_t + v_t, & v_t \sim (0, V)
\end{array}
\tag{107}
$$

where $\{\beta_0\}$, $\{w_t : t = 1, \dots, N\}$, and $\{v_t : t = 1, \dots, N\}$ are mutually independent collections of independent random variables; where the system equation is true for $t = 1, \dots, N$ and the observation equation is true for all $Y_t \in \boldsymbol{\mathcal{F}}_N$, i.e. for $t = 1, \dots, N$; where β_t for $t = 0, \dots, N$ are scalars and $Y_t \in \boldsymbol{\mathcal{F}}_N$ are scalars; and where $|\phi| < 1$ and $|\eta| < 1$. Defining a new sequence of smoother estimates as $\hat{\boldsymbol{\beta}}_{0:N} \equiv \{\hat{\beta}_{t|N} : t = 0, \dots, N\}$ that satisfy the following system of state estimating equations similar to (92)

$$
\begin{aligned}
\boldsymbol{M}_N \hat{\beta}_{N:0|N} = \boldsymbol{Y}^*_{N:0|N} \text{ or } \hat{\boldsymbol{\beta}}_{N:0|N} &= \boldsymbol{M}_N^{-1} \boldsymbol{Y}^*_{N:0} \\
\hat{\boldsymbol{\beta}}_{N:0|N} &\equiv \left(\hat{\beta}_{t|N} : t = N, \dots, 0\right)'
\end{aligned}
$$

such that the distribution for the associated collection of smoother residuals defined as $\tilde{\boldsymbol{\beta}}_{0:N} \equiv \{\tilde{\beta}_{t|N} \equiv \beta_t - \hat{\beta}_{t|N} : t = 0, \dots N\}$ satisfies

$$
\begin{aligned}
\boldsymbol{M}_N \tilde{\boldsymbol{\beta}}_{N:0|N} \sim (\mathbf{0}, \boldsymbol{M}_N) \text{ or } \tilde{\boldsymbol{\beta}}_{N:0|N} &\sim \left(\mathbf{0}, \boldsymbol{M}_N^{-1}\right) \\
\tilde{\boldsymbol{\beta}}_{N:0|N} &\equiv (\tilde{\beta}_{t|N} : t = N, \dots, 0)' .
\end{aligned}
$$

Hence the results of this section are applicable to the smoother residuals in $\tilde{\boldsymbol{\beta}}_{0:N}$ associated with the linear state space model defined in (107).

3.4.1 Missing Observations

In this section, the results of the previous section are generalized for the case where some observations are unavailable, both in the past and in the future given a reference time point. When there are no missing observations, then the results of this section reduce to the results of the previous section where all observations are available. Initially the Linear Gaussian State Space model, as defined in (91), is assumed true. Denote the available observations as $\boldsymbol{\mathcal{F}}_{N^*}$ and denote the available observation index as N^*

$$
\begin{aligned}
N^* &\equiv \{t \in [1, \dots, N] : Y_t \text{ is available } \} \\
\boldsymbol{\mathcal{F}}_{N^*} &\equiv \{Y_t : t \in N^*\} .
\end{aligned}
$$

The conditional distribution of $\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_{N^*}$ is a multivariate Gaussian distribution since the distribution of $(\beta_{0:N}, \boldsymbol{\mathcal{F}}_{N^*})$ is a multivariate Gaussian distribution. Consequently, finding the mode of the posterior distribution for $\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_{N^*}$ is equivalent to finding the mean of the posterior distribution for $\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_{N^*}$. The posterior distribution for $\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_{N^*}$ is given by

$$f\left(\boldsymbol{\beta}_{0:N}|\boldsymbol{\mathcal{F}}_{N^*}\right) = \left[\prod_{t\in N^*} f\left(Y_t|\beta_t\right)\right]\left[\prod_{t=1}^{N} f\left(\beta_t|\beta_{t-1}\right)\right] f\left(\beta_0\right)/f\left(\boldsymbol{\mathcal{F}}_{N^*}\right) \ .$$

The mode (and mean) $\boldsymbol{\beta}^k_{N^*} = \{\beta_{0|N^*}, \ldots, \beta_{N|N^*}\}$ of the posterior distribution can be found by maximizing the log likelihood using

$$\begin{aligned} \mathbf{0}_{(N+1)} &= \boldsymbol{\nabla} \log f\left(\beta_{0:N}|\, \boldsymbol{\mathcal{F}}_{N^*}\right)|_{\beta^k_{N^*}} \\ \boldsymbol{\nabla} &\equiv \left(\frac{\partial}{\partial\beta_0}, \ldots, \frac{\partial}{\partial\beta_N}\right)' \ . \end{aligned}$$

The resulting system of state estimating equations can be written as

$$\begin{bmatrix} A_{N|N^*} & -C & & & \\ -C & B_{N-1|N^*} & -C & & \\ & & \ddots & & \\ & & -C & B_{1|N^*} & -C \\ & & & -C & D \end{bmatrix} \begin{pmatrix} \beta_{N|N^*} \\ \beta_{N-1|N^*} \\ \vdots \\ \beta_{1|N^*} \\ \beta_{0|N^*} \end{pmatrix} = \begin{pmatrix} \frac{\eta}{V} Y_{N|N^*} \\ \frac{\eta}{V} Y_{N-1|N^*} \\ \vdots \\ \frac{\eta}{V} Y_{1|N^*} \\ \frac{b_0}{W_0} \end{pmatrix}$$

where

$$\begin{aligned} A_{N|N^*} &= \begin{cases} \frac{1}{W} + \frac{\eta^2}{V} & : N \in N^* \\ \frac{1}{W} & : N \notin N^* \end{cases} \\ B_{t|N^*} &= \begin{cases} \frac{1+\phi^2}{W} + \frac{\eta^2}{V} & : t \in N^* \\ \frac{1+\phi^2}{W} & : t \notin N^* \end{cases} \quad t = 1, \ldots, N-1 \\ C &= \frac{\phi}{W} \\ D &= \frac{1}{W_0} + \frac{\phi^2}{W} \\ Y_{t|N^*} &= \begin{cases} Y_t & : t \in N^* \\ 0 & : t \notin N^* \end{cases} \end{aligned}$$

or in matrix notation as

$$\boldsymbol{M}_{N^*}\boldsymbol{\beta}^k_{N:0|N^*} = \boldsymbol{Y}^*_{N:0|N^*} \tag{108}$$

where $\boldsymbol{\beta}^{k}_{N:0|N^*} \equiv (\beta_{N|N^*}, \ldots, \beta_{0|N^*})'$ is a vector of the state space smoother estimates for the state vector $\boldsymbol{\beta}_{N:0} \equiv (\beta_N, \ldots, \beta_0)'$ given the available observations in $\boldsymbol{\mathcal{F}}_{N^*}$.

It is easy to show that $\boldsymbol{M}_{N^*}$ is positive definite and invertible. Analyzing the system of equations associated with $\boldsymbol{\nabla} \log f(\beta_{0:N}|\boldsymbol{\mathcal{F}}_{N^*})$, when the Linear Gaussian State Space model (91) is true with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*}$, and defining the vector of smoother residuals as $\tilde{\boldsymbol{\beta}}_{N:0|N^*} \equiv (\tilde{\beta}_{t|N^*} \equiv \beta_t - \beta_{t|N^*} : t = N, \ldots, 0)'$, shows

$$\boldsymbol{M}_{N^*}\tilde{\boldsymbol{\beta}}_{N:0|N^*} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}_{N^*}\right) \text{ or } \tilde{\boldsymbol{\beta}}_{N:0|N^*} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}^{-1}_{N^*}\right) \ .$$

Similar to previous results in the previous section, entries in $\boldsymbol{M}^{-1}_{N^*}$ are calculated to find the precision values $P_{t|N^*} = \mathrm{Var}\ \tilde{\beta}_{t|N^*}$, for $t = 0, \ldots, N$.

Lemma 3.7. *Given the linear Gaussian state space model defined in (91) with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*}$*

$$\begin{aligned}
Var\ \tilde{\beta}_{0|N^*} &= \left(D - \frac{C^2}{G_{N|N^*}}\right)^{-1} \\
Cov\left(\tilde{\beta}_{0|N^*}, \tilde{\beta}_{t|N^*}\right) &= \frac{C}{G_{N-t+1|N^*}} \times \cdots \times \frac{C}{G_{N|N^*}}\, Var\ \tilde{\beta}_{0|N^*},\ t = 1, \ldots, N \\
G_{j|N^*} &\equiv \begin{cases} A_{N|N^*} & : j = 1 \\ B_{N-j+1|N^*} - \frac{C^2}{G_{j-1|N^*}} & : 1 < j \leq N \end{cases} .
\end{aligned}$$

Proof: The result is proven by using Gaussian elimination to solve $\boldsymbol{M}_{N^*}\boldsymbol{X}_N = \boldsymbol{e}_{N+1}$ where $\boldsymbol{e}_{N+1} = (0, \ldots, 0, 1)'$. The Gaussian elimination of $\boldsymbol{M}_{N^*}$ proceeds by eliminating the lower diagonal starting from the left and then by eliminating the upper diagonal starting from the right. ■

Lemma 3.8. *Given the linear Gaussian state space model defined in (91) with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*}$*

$$\begin{aligned}
Var\ \tilde{\beta}_{N|N^*} &= \left(A_{N|N^*} - \frac{C^2}{G^*_{N|N^*}}\right)^{-1} \\
Cov\left(\tilde{\beta}_{t|N^*}, \tilde{\beta}_{N|N^*}\right) &= \frac{C}{G^*_{t+1|N^*}} \times \cdots \times \frac{C}{G^*_{N|N^*}}\, Var\ \tilde{\beta}_{N|N^*},\ t = 0, \ldots, N-1 \\
G^*_{j|N^*} &\equiv \begin{cases} D & : j = 1 \\ B_{j-1|N^*} - \frac{C^2}{G^*_{j-1|N^*}} & : 1 < j \leq N \end{cases} .
\end{aligned}$$

Proof: The result is proven by using Gaussian elimination to solve $\boldsymbol{M}_{N^*}\boldsymbol{X}_N = \boldsymbol{e}_1$ where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$. The Gaussian elimination of $\boldsymbol{M}_{N^*}$ proceeds by eliminating the upper diagonal starting from the right and then by eliminating the lower diagonal starting from the left. ■

Lemma 3.9. *Given the linear Gaussian state space model defined in (91) with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*}$ then for $t = 1, \ldots, N-1$, $t_1 = 0, \ldots, t-1$, and $t_2 = t+1, \ldots, N$*

$$\begin{aligned}
Var\ \tilde{\beta}_{t|N^*} &= \left(G_{N-t+1|N^*} - \frac{C^2}{G^*_{t|N^*}}\right)^{-1} \\
Cov\left(\tilde{\beta}_{t_1|N^*}, \tilde{\beta}_{t|N^*}\right) &= \frac{C}{G^*_{t_1+1|N^*}} \times \cdots \times \frac{C}{G^*_{t|N^*}}\, Var\ \tilde{\beta}_{t|N^*} \\
Cov\left(\tilde{\beta}_{t|N^*}, \tilde{\beta}_{t_2|N^*}\right) &= \frac{C}{G_{N-t_2+1|N^*}} \times \cdots \times \frac{C}{G_{N-t|N^*}}\, Var\ \tilde{\beta}_{t|N^*}
\end{aligned}$$

where $G_{j|N^*}$ and $G^*_{j|N^*}$ have been previously defined in Lemmas 3.7 and 3.8.

Proof: Given a fixed $t \in [1, \ldots, N-1]$, the result is proven by using Gaussian elimination to solve $\boldsymbol{M}_{N^*}\boldsymbol{X}_N = \boldsymbol{e}_{N-t+1}$ where $\boldsymbol{e}_{N-t+1}$ is a vector consisting of $N+1$ zeros except for a one in element number $N-t+1$. The Gaussian elimination of $\boldsymbol{M}_{N^*}$ proceeds by eliminating $N-t$ elements in the lower diagonal starting from the left and then by eliminating t elements in the upper diagonal starting from the right. The remainder of the elements in the upper and lower diagonals are then eliminated. ■

As expected, the missing observation precisions are bounded by the two cases where all observations Y_t are available for $t = 1, \ldots, N$, and where no observations Y_t are available for $t = 1, \ldots, N$.

Proposition 3.5. *Given the linear Gaussian state space model defined in (91) with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*}$*

$$Var\ \tilde{\beta}_{t|N} \leq Var\ \tilde{\beta}_{t|N^*} \leq Var\ \tilde{\beta}_{t|N^0},\ t = 0, \ldots, N$$

where $Var\ \tilde{\beta}_{t|N}, t = 0, \ldots, N$ are the precision values associated with

$$\boldsymbol{\mathcal{F}}_N = \{Y_t \text{ available for } t = 1, \ldots, N\}$$

where $Var\ \tilde{\beta}_{t|N^0}, t = 0, \ldots, N$ are the precision values associated with

$$\boldsymbol{\mathcal{F}}_{N^0} = \{Y_t \text{ unavailable for } t = 1, \ldots, N\} = \emptyset = N^0$$

and where

$$
\begin{aligned}
G_{j|N^0} &= A_0 = \frac{1}{W},\ j = 1, \ldots, N \\
G^*_{j|N^0} &= \begin{cases} D & : j = 1 \\ B_0 - \frac{C^2}{G^*_{j-1|N^0}} & : j > 1 \end{cases} \\
B_0 &= A_0 + \frac{C^2}{A_0} = \frac{1+\phi^2}{W}\ .
\end{aligned}
$$

Proof: With regards to the lower bounds, $\boldsymbol{M}_N = \boldsymbol{M}_{N^*} + \boldsymbol{M}_{(1)}$ where $\boldsymbol{M}_{N^*} > \mathbf{0}$ and where

$$
\boldsymbol{M}_{(1)} = \begin{bmatrix} A - A_{N|N^*} & & & & \\ & B - B_{N-1|N^*} & & & \\ & & \ddots & & \\ & & & B - B_{1|N^*} & \\ & & & & 0 \end{bmatrix} \geq \mathbf{0}\ .
$$

Hence $\boldsymbol{M}_{N^*} \leq \boldsymbol{M}_N$ implies $\boldsymbol{M}^{-1}_{N^*} \geq \boldsymbol{M}^{-1}_N$ proving the result for the lower bounds.

With regards to the upper bounds, $\boldsymbol{M}_{N^*} \geq \boldsymbol{M}_{(0)}$ where

$$
\boldsymbol{M}_{(0)} = \begin{bmatrix} A_0 & -C & & & \\ -C & B_0 & -C & & \\ & & \ddots & & \\ & & -C & B_0 & -C \\ & & & -C & D \end{bmatrix} > \mathbf{0}\ .
$$

By direct examination, $\boldsymbol{M}_{(0)} = \boldsymbol{M}_{N^0}$ associated with $\boldsymbol{\mathcal{F}}_{N^0}$. Hence $\boldsymbol{M}^{-1}_{N^*} \leq \boldsymbol{M}^{-1}_{N^0}$ proving the result for the upper bounds. ■

The asymptotic analysis of the precision values as $N \to \infty$ is shown for two cases, where there is a finite number of available observations, or where there is a finite number of missing observations. The first case includes Kalman prediction of those states beyond the last available observation.

Proposition 3.6. *Given the linear Gaussian state space model defined in (91) with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*} = \boldsymbol{\mathcal{F}}_{n^*}$ for $N > n$ such that $Y_n \in \boldsymbol{\mathcal{F}}_{n^*}$ denotes the* *last available observation, then as* $N \to \infty$

$$\begin{aligned} Var\ \tilde{\beta}_{0|N^*} &= \left(D - \frac{C^2}{G_{n|n^*}}\right)^{-1} \\ Var\ \tilde{\beta}_{t|N^*} &\to \left(G_{n-t+1|n^*} - \frac{C^2}{G^-_{t|n^*}}\right)^{-1} \quad \textit{for fixed } t \in [1, \ldots, n] \\ Var\ \tilde{\beta}_{t|N^*} &\to \left(\frac{1}{W} - \frac{C^2}{G^-_{t|n^*}}\right)^{-1} \quad \textit{for fixed } t \in (n, \ldots, \infty) \end{aligned}$$

where

$$G^-_{j|n^*} \equiv \begin{cases} D & : j = 1 \\ B_{j-1|n*} - \frac{C^2}{G^-_{j-1|n^*}} & : 1 < j \le n \\ B_0 - \frac{C^2}{G^-_{j-1|n^*}} & : n < j \end{cases} .$$

Proof: None of the observations are available for $t \in [n+1, \ldots, N]$. The definitions of $G_{j|N^*}$ and $G^*_{j|N^*}$ are used to show

$$\begin{aligned} G_{N-t+1|N^*} &= \begin{cases} \frac{1}{W} & : n < t \le N \\ G_{n-t+1|n^*} & : 1 \le t \le n \end{cases} \\ G^*_{t|N^*} &= G^-_{t|n^*},\ 1 \le t \le N . \end{aligned}$$

The previous display proves the result for $t \in [0, \ldots, N]$ since

$$\begin{aligned} \text{Var}\ \tilde{\beta}_{0|N^*} &= \left(D - \frac{C^2}{G_{N|N^*}}\right)^{-1} \\ \text{Var}\ \tilde{\beta}_{t|N^*} &= \left(G_{n-t+1|N^*} - \frac{C^2}{G^*_{t|N^*}}\right)^{-1},\ 1 \le t < N . \end{aligned}$$

Allowing $N \to \infty$ completes the proof. ■

Proposition 3.7. *Given the linear Gaussian state space model defined in (91) with* $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*} \supset \boldsymbol{\mathcal{F}}_{n^*}$ *where* $\boldsymbol{\mathcal{F}}_{n^*}$ *contains the available observations for* $t \in [1, \ldots, n]$ *with* $N > n$ *such that* $Y_n \notin \boldsymbol{\mathcal{F}}_{n^*}$ *denotes the last missing* *observation, then as* $N \to \infty$

$$\mathit{Var}\ \tilde{\beta}_{0|N^*} \to \left(D - \frac{C^2}{G^{\infty}_{n+1|n^*}} \right)^{-1}$$

$$\mathit{Var}\ \tilde{\beta}_{t|N^*} \to \left(G^{\infty}_{n-t+2|n^*} - \frac{C^2}{G^{+}_{t|n^*}} \right)^{-1} \quad \textit{for fixed } t \in [1, \ldots, n]$$

$$\mathit{Var}\ \tilde{\beta}_{t|N^*} \to \left(G_{\infty} - \frac{C^2}{G^{+}_{t|n^*}} \right)^{-1} \quad \textit{for fixed } t \in (n, \ldots, \infty)$$

where

$$G^{+}_{j|n^*} \equiv \begin{cases} D & : j = 1 \\ B_{j-1|n*} - \frac{C^2}{G^{+}_{j-1|n^*}} & : 1 < j \le n \\ B - \frac{C^2}{G^{+}_{j-1|n^*}} & : n < j \end{cases}$$

$$G^{\infty}_{j|n^*} \equiv \begin{cases} G_{\infty} & : j = 1 \\ B_0 - \frac{C^2}{G^{\infty}_{1|n^*}} & : j = 2 \\ B_{n-j+2|n^*} - \frac{C^2}{G^{\infty}_{j-1|n^*}} & : 2 < j \le n+1 \end{cases} .$$

Proof: All of the observations are available for $t \in [n+1, \ldots, N]$. The definition of $G_{j|N^*}$ with $j \in [1, \ldots, N-n]$ is used to define $G_{N-t+1|N^*}$ with $t \in [n+1, \ldots, N]$ in order to show as $N \to \infty$

$$G_{N-t+1|N^*} \to G^{\infty}_{1|n^*} \text{ for } t \in [n+1, \ldots, \infty) \ .$$

The continuity of $G_{N-t+1|N^*}$ as a function of $G_{N-n|N^*}$ with $t \in [1, \ldots, n]$ is used to show as $N \to \infty$

$$G_{N-t+1|N^*} \to G^{\infty}_{n-t+2|n^*} \text{ for } t \in [1, \ldots, n] \ .$$

Note that $G^{*}_{j|N^*} = G^{+}_{j|n^*}$ for $1 \le j \le N$. Hence, the result is proven by starting with the following equations and allowing $N \to \infty$

$$\text{Var } \tilde{\beta}_{0|N^*} = \left(D - \frac{C^2}{G_{N|N^*}} \right)^{-1}$$

$$\text{Var } \tilde{\beta}_{t|N^*} = \left(G_{N-t+1|N^*} - \frac{C^2}{G^{*}_{t|N^*}} \right)^{-1}, \ 1 \le t \le N \ . \ \blacksquare$$

The following corollary checks the results of Proposition 3.6 against the Kalman prediction method.

Corollary 3.6. *Given the linear Gaussian state space model defined in (91) with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*} = \boldsymbol{\mathcal{F}}_{n^*}$ for $N > n$ such that $Y_n \in \boldsymbol{\mathcal{F}}_{n^*}$ denotes the last available observation, then for $t \in [n, \ldots, N-1]$*

$$Var\ \tilde{\beta}_{t+1|N^*} = \phi^2\, Var\ \tilde{\beta}_{t|N^*} + W\ .$$

Proof: Applying the equation for Var $\tilde{\beta}_{t|N^*}$, applying the following identity for $t \in [0, \ldots, N-1]$,

$$\frac{\text{Var}\ \tilde{\beta}_{t+1|N^*}}{\text{Var}\ \tilde{\beta}_{t|N^*}} = \frac{G^*_{t+1|N^*}}{G_{N-t|N^*}}$$

and using a little algebra shows that proving the result is equivalent to proving the following display for $t \in [n, \ldots, N-1]$

$$\left(\phi^2 + W G^*_{t+1|N^*}\right) G_{N-t|N^*} - G^*_{t+1|N^*} = \frac{\phi^2}{W}\ .$$

Proposition 3.6 shows that $G_{N-t|N^*} = 1/W$ for $t \in [n, \ldots, N-1]$. Hence the previous display and the result are proven. ■

Remark 3.3. With respect to the linear state space model as defined in (107) where the Gaussian assumption has been removed and with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*}$, define a new collection of smoother estimates as $\hat{\boldsymbol{\beta}}_{N^*} \equiv \{\hat{\beta}_{t|N^*}, t = 0, \ldots, N\}$ that satisfy the following system of state estimating equations similar to (108)

$$\begin{gathered}\boldsymbol{M}_{N^*}\hat{\beta}_{N:0|N^*} = \boldsymbol{Y}^*_{N:0|N^*} \text{ or } \hat{\boldsymbol{\beta}}_{N:0|N^*} = \boldsymbol{M}^{-1}_{N^*}\boldsymbol{Y}^*_{N:0|N^*}\\ \hat{\boldsymbol{\beta}}_{N:0|N^*} \equiv \left(\hat{\beta}_{t|N^*} : t = N, \ldots, 0\right)'\ .\end{gathered}$$

The distribution for the associated collection of smoother residuals defined as $\tilde{\boldsymbol{\beta}}_{N^*} \equiv \{\tilde{\beta}_{t|N^*} \equiv \beta_t - \hat{\beta}_{t|N^*}, t = 0, \ldots, N\}$ satisfies

$$\begin{gathered}\boldsymbol{M}_{N^*}\tilde{\boldsymbol{\beta}}_{N:0|N^*} \sim (\boldsymbol{0}, \boldsymbol{M}_{N^*}) \text{ or } \tilde{\boldsymbol{\beta}}_{N:0|N^*} \sim \left(\boldsymbol{0}, \boldsymbol{M}^{-1}_{N^*}\right)\\ \tilde{\boldsymbol{\beta}}_{N:0|N^*} \equiv \left(\tilde{\beta}_{t|N^*} : t = N, \ldots, 0\right)'\ .\end{gathered}$$

Hence the results of this section are applicable to the smoother residuals in $\tilde{\boldsymbol{\beta}}_{N^*}$ associated with the linear state space model defined in (107) with $\boldsymbol{\mathcal{F}}_N = \boldsymbol{\mathcal{F}}_{N^*}$.

3.5 Partial State Space Smoother

This section introduces the partial state space smoother that generates a collection of partial smoother estimates of each state at time t that depends on only a finite number of past, current, and future observations relative to time t. The number of operations needed by the partial state space smoother is fewer than the number of operations needed by the complete state space smoother, at the price of larger precisions for the partial smoother estimates relative to the precisions for the complete smoother estimates.

In order to motivate the partial state space smoother, consider the collection of complete smoother estimates $\boldsymbol{\beta}^k_{N:0|N}$ under the linear Gaussian state space model (91) that satisfies the tridiagonal system of state estimating equations from (92) as follows

$$\boldsymbol{M}_N \boldsymbol{\beta}^k_{N:0|N} = \boldsymbol{Y}^*_{N:0|N}, \; \boldsymbol{\beta}^k_{N:0|N} = \left(\beta_{N|N}, \ldots, \beta_{0|N}\right)'$$

$$\boldsymbol{M}_N = \begin{bmatrix} A & -C & & & \\ -C & B & -C & & \\ & & \ddots & & \\ & & -C & B & -C \\ & & & -C & D \end{bmatrix} \in \mathbb{R}^{N+1 \times N+1} \, .$$

As noted in Remark 3.1, this system of state estimating equations can be solved by the likelihood smoother that uses Gaussian elimination to remove the upper diagonal and then the lower diagonal of $\boldsymbol{M}_N$. When new observations become available then the lower diagonal of $\boldsymbol{M}_N$ needs to be removed again. As N gets large, the number of operations needed by the complete state space smoother also gets large. The power (i.e. minimum precision) of the complete smoother estimates comes from the tridiagonal structure of $\boldsymbol{M}_N$. The cost of this power is the number of operations needed to diagonalize $\boldsymbol{M}_N$. One way to decrease the number of operations conceptually is to decrease the number of backward links in the lower diagonal of $\boldsymbol{M}_N$ such that each of the resulting partial smoother estimates only rely on a subset of the N observations. The penalty for removing backward links in $\boldsymbol{M}_N$ shows up in the power (by an increase in the precision) of the resulting partial smoother estimates.

Sections 3.5.1 through 3.5.3 introduce a partial smoother that solves a system of state estimating equations with all or part of the lower diagonal removed. Section 3.5.4 describes another partial smoother that solves a system of state estimating equations different from both the generalized partial state space smoother of section 3.5.2 and the complete state space smoother.

3.5.1 A Simple Partial Smoother

As the first example of a partial state space smoother given the linear Gaussian state space model (91), consider a collection of new partial smoothers $\hat{\boldsymbol{\beta}}^l_{0:N} \equiv \{\hat{\beta}^l_{t|t} : t = 0, \ldots, N\}$ that satisfy the following new system of state estimating equations

$$-\frac{1}{W}\left(\hat{\beta}^l_{t|t} - \phi\hat{\beta}^l_{t-1|t-1}\right) + \frac{\eta}{V}\left(Y_t - \eta\hat{\beta}^l_{t|t}\right) = 0, \; t = N, \ldots, 1$$
$$-\frac{1}{W_0}\left(\hat{\beta}^l_{0|0} - b_0\right) = 0$$

that is written in matrix notation as

$$\begin{bmatrix} A & -C & & & \\ & A & -C & & \\ & & \ddots & & \\ & & & A & -C \\ & & & & D_0 \end{bmatrix} \begin{pmatrix} \hat{\beta}^l_{N|N} \\ \hat{\beta}^l_{N-1|N-1} \\ \vdots \\ \hat{\beta}^l_{1|1} \\ \hat{\beta}^l_{0|0} \end{pmatrix} = \begin{pmatrix} \frac{\eta}{V}Y_N \\ \frac{\eta}{V}Y_{N-1} \\ \vdots \\ \frac{\eta}{V}Y_1 \\ \frac{b_0}{W_0} \end{pmatrix}$$
$$A = \frac{1}{W} + \frac{\eta^2}{V}, \; C = \frac{\phi}{W}, \; D_0 = \frac{1}{W_0}$$

that is represented in matrix symbology as

$$\boldsymbol{U}_l \hat{\boldsymbol{\beta}}^l_{N:0} = \boldsymbol{Y}^*_{N:0} \tag{109}$$

and that is different from the system of state estimating equations associated with the complete state space smoother (92) since the principle lower diagonal is $\mathbf{0}$. It is easy to see that $\boldsymbol{U}_l$ is upper diagonal and invertible such that

$$\hat{\boldsymbol{\beta}}^l_{N:0} = \boldsymbol{U}_l^{-1}\boldsymbol{Y}^*_{N:0}$$

and such that each of the partial smoothers can be found recursively using

$$\hat{\beta}^l_{0|0} = b_0$$
$$\hat{\beta}^l_{t|t} = A^{-1}\left(\frac{\eta}{V}Y_t + C\hat{\beta}^l_{t-1|t-1}\right), \; t = 1, \ldots, N \; .$$

Hence each partial smoother $\hat{\beta}^l_{t|t}$ for $t \in \{1, \ldots, N\}$ depends linearly on only the observations Y_1 through Y_t.

Substituting the states $\boldsymbol{\beta}_{N:0}$ for the partial smoothers $\hat{\boldsymbol{\beta}}^{l}_{N:0}$ in (109) and applying the linear Gaussian state space model (91) results in

$$\boldsymbol{U}_l \boldsymbol{\beta}_{N:0} - \boldsymbol{Y}^*_{N:0} = \begin{pmatrix} \frac{1}{W} w_N - \frac{\eta}{V} v_N \\ \vdots \\ \frac{1}{W} w_1 - \frac{\eta}{V} v_1 \\ \frac{1}{W_0}\left(\beta_0 - b_0\right) \end{pmatrix} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{D}_l\right)$$
$$\boldsymbol{D}_l \equiv \text{Diagonal}\begin{pmatrix} A & \dots & A & D_0 \end{pmatrix}$$

where $\boldsymbol{D}_l$ is a diagonal matrix. Define the collection of partial smoother residuals as $\tilde{\boldsymbol{\beta}}^{l}_{0:N} \equiv \{\tilde{\beta}^{l}_{t|t} \equiv \beta_t - \hat{\beta}^{l}_{t|t} : t = 0, \dots, N\}$. Hence the partial smoother residuals $\tilde{\boldsymbol{\beta}}^{l}_{0:N}$ satisfy the following relationship

$$\boldsymbol{U}_l \tilde{\boldsymbol{\beta}}^{l}_{N:0} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{D}_l\right) \text{ or } \tilde{\boldsymbol{\beta}}^{l}_{N:0} \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{M}_l^{-1}\right)$$
$$\tilde{\boldsymbol{\beta}}^{l}_{N:0} \equiv \left(\tilde{\beta}^{l}_{t|t} : t = N, \dots, 0\right)'$$
$$\boldsymbol{M}_l^{-1} \equiv \left(\boldsymbol{U}_l\right)^{-1} \boldsymbol{D}_l \left(\boldsymbol{U}_l'\right)^{-1} \ .$$

Using matrix multiplication shows that

$$\begin{aligned} \boldsymbol{M}_l &= \boldsymbol{U}_l' \boldsymbol{D}_l^{-1} \boldsymbol{U}_l \\ &= \begin{bmatrix} A & -C & & & \\ -C & A + \frac{C^2}{A} & -C & & \\ & & \ddots & & \\ & & -C & A + \frac{C^2}{A} & -C \\ & & & -C & D_0 + \frac{C^2}{A} \end{bmatrix} . \end{aligned}$$

The previous display leads directly to a lower bounds on Var $\tilde{\boldsymbol{\beta}}^{l}_{N:0}$ and to simple formulas for each Var $\tilde{\beta}_{t|t}, t = 0, \dots, N$.

Proposition 3.8. *Given the linear Gaussian state space model in (91) then*

$$\mathit{Var}\ \tilde{\boldsymbol{\beta}}^{l}_{N:0} \geq \mathit{Var}\ \tilde{\boldsymbol{\beta}}_{N:0|N}$$

where equality exists if and only if $\eta = 0$.

Proof: Simple algebra shows that

$$\frac{C^2}{A} < \frac{\phi^2}{W} \text{ for } \eta \neq 0, \ \frac{C^2}{A} = \frac{\phi^2}{W} \text{ for } \eta = 0\ .$$

Hence the result is proven since

$$\boldsymbol{M}_l = (\text{Var } \tilde{\boldsymbol{\beta}}^l_{N:0})^{-1} \leq \boldsymbol{M}_N = (\text{Var } \tilde{\boldsymbol{\beta}}_{N:0|N})^{-1}$$

implies Var $\tilde{\boldsymbol{\beta}}^l_{N:0} \geq$ Var $\tilde{\boldsymbol{\beta}}_{N:0|N}$ with equality if and only if $\eta = 0$. ■

Lemma 3.10. *Given the linear Gaussian state space model in (91) then*

$$\begin{aligned} \mathit{Var}\ \tilde{\beta}^l_{0|0} &= W_0 \\ \mathit{Var}\ \tilde{\beta}^l_{t|t} &= \left(A - \frac{C^2}{G^*_{t|l}}\right)^{-1},\ t = 1, \ldots, N\ . \end{aligned}$$

where

$$G^*_{j|l} = \begin{cases} D_0 + \frac{C^2}{A} & : j = 1 \\ A + \frac{C^2}{A} - \frac{C^2}{G^*_{j-1|l}} & : j > 1 \end{cases}.$$

Proof: Given a fixed $t \in [0, \ldots, N]$, Gaussian elimination of $\boldsymbol{M}_l \boldsymbol{X}_N = \boldsymbol{e}_{N-t+1}$ is used to show that

$$\begin{aligned} \text{Var } \tilde{\beta}^l_{0|0} &= \left(D_0 + \frac{C^2}{A} - \frac{C^2}{G_{N|l}}\right)^{-1} \\ \text{Var } \tilde{\beta}^l_{t|t} &= \left(G_{N-t+1|l} - \frac{C^2}{G^*_{t|l}}\right)^{-1},\ t = 1, \ldots, N-1 \\ \text{Var } \tilde{\beta}^l_{N|N} &= \left(A - \frac{C^2}{G^*_{N|l}}\right)^{-1} \end{aligned}$$

where

$$G_{j|l} = \begin{cases} A & : j = 1 \\ A + \frac{C^2}{A} - \frac{C^2}{G_{j-1|l}} & : j > 1 \end{cases}.$$

Noting that $G_{j|l} = A$ for $j = 1, \ldots, N$ proves the result. ■

Bounds for each Var $\tilde{\beta}^l_{t|t}$, $t \in [1, \ldots, N]$, are also found using the properties of $G^*_{j|l}$.

Lemma 3.11. *The properties of $G^*_{j|l}$ include the following*

$$\textit{If } G^*_{1|l} < A \textit{ then } \frac{C^2}{A} < G^*_{j|l} < G^*_{j+1|l} < A,\ j = 2, \ldots \tag{110}$$

$$\textit{If } G^*_{1|l} > A \textit{ then } A < G^*_{j+1|l} < G^*_{j|l} < A + \frac{C^2}{A},\ j = 2, \ldots \tag{111}$$

$$G^*_{j|l} \to A \textit{ as } j \to \infty\ . \tag{112}$$

Proof: The fact that $C^2/A < G^*_{1|l}$ is used to show $C^2/A < G^*_{2|l}$. Induction is used to show the general result that for $j = 2, \ldots$

$$\frac{C^2}{A} < G^*_{j|l} < A + \frac{C^2}{A} \ .$$

Simple algebra is used to show for $j = 2, \ldots,$

$$\begin{aligned}
\text{If } G^*_{j-1|l} < G^*_{j|l} &\text{ then } G^*_{j|l} < G^*_{j+1|l} \\
\text{If } G^*_{j-1|l} > G^*_{j|l} &\text{ then } G^*_{j|l} > G^*_{j+1|l} \\
\text{If } G^*_{j|l} < A &\text{ then } G^*_{j+1|l} < A \\
\text{If } G^*_{j|l} > A &\text{ then } G^*_{j+1|l} > A \ .
\end{aligned}$$

Algebraic analysis also shows that $G^*_{1|l} \leq A$ is equivalent to $G^*_{1|l} \leq G^*_{2|l}$. Induction utilizing the inequalities in the previous display proves the result for (110). A similar analysis proves the result for (111). Results (110) and (111) show that $G^*_{j|l} \to G^*_{\infty|l}$ as $j \to \infty$. Hence the identity

$$G^*_{\infty|l} = A + \frac{C^2}{A} - \frac{C^2}{G^*_{\infty|l}}$$

has two solutions: $G^*_{\infty|l} = A, C^2/A$. The first solution, $G^*_{\infty|l} = A$, is the only solution that satisfies the previous results (110) and (111). Hence the result (112) is proven. ∎

Proposition 3.9. *Given the linear Gaussian state space model in (91) then each Var $\tilde{\beta}^l_{t|t}$ for $t = 0, \ldots, N$ is bounded as follows*

$$Var \ \tilde{\beta}^l_{0|0} = W_0$$

If $G^*_{1|l} < A$ then for $t = 1, \ldots, N$

$$\left(A - \frac{C^2}{A}\right)^{-1} < \tilde{\beta}^l_{t|t} < \left(A - \frac{C^2}{G^*_{1|l}}\right)^{-1}$$

Else if $G^*_{1|l} > A$ then for $t = 1, \ldots, N$

$$\left(A - \frac{C^2}{G^*_{1|l}}\right)^{-1} < \tilde{\beta}^l_{t|t} < \left(A - \frac{C^2}{A}\right)^{-1} \ .$$

Var $\tilde{\beta}^l_{N|N}$ converges to a limit as $N \to \infty$

$$Var\ \tilde{\beta}^l_{N|N} \to \left(A - \frac{C^2}{A}\right)^{-1} .$$

■

The following corollary verifies that the precisions of the Kalman filter estimates are smaller than the precisions of the partial smoother estimates. The next section shows that the precisions of the state space smoother estimates are also smaller than the precisions of the partial smoother estimates since additional observations are used to calculate the state space smoother estimate versus the partial smoother estimate of each state.

Corollary 3.7. *Given the linear Gaussian state space model in (91) then the precisions of the Kalman filter estimates $P_{t|t} = Var\ \tilde{\beta}_{t|t}$ are smaller than the precisions of the partial smoother estimates $P^l_{t|t} = Var\ \tilde{\beta}^l_{t|t}$*

$$\begin{aligned} Var\ \tilde{\beta}_{t|t} &< Var\ \tilde{\beta}^l_{t|t},\ \textit{for}\ t \in [1, \ldots, \infty) \\ \lim_{N\to\infty} Var\ \tilde{\beta}_{N|N} &< \lim_{N\to\infty} Var\ \tilde{\beta}^l_{N|N} . \end{aligned}$$

Proof: With regards to the first result, direct examination shows that $G^*_1 < G^*_{1|l}$. Hence $G^*_2 < G^*_{2|l}$ by direct calculation and $G^*_j < G^*_{j|l}$ for $j = 3, \ldots$ by induction. The first result follows by using the equations for Var $\tilde{\beta}_{t|t}$ and Var $\tilde{\beta}^l_{t|t}$.

The second result follows from the equation for Var $\tilde{\beta}_{N|N}$ and from the convergence of $G^*_N \to G_\infty$ and $G^*_{N|l} \to A$ as $N \to \infty$ such that $G_\infty > A$. ■

In order to further compare these precisions, the ratio of the precision for the Kalman filter estimate Var $\tilde{\beta}_{N|N}$ versus the precision for the partial smoother estimate Var $\tilde{\beta}^l_{N|N}$ is examined as $N \to \infty$

$$\begin{aligned} \lim_{N\to\infty} \frac{\text{Var}\ \tilde{\beta}_{N|N}}{\text{Var}\ \tilde{\beta}^l_{N|N}} &= \frac{A - \frac{C^2}{A}}{A - \frac{C^2}{G_\infty}} = \frac{A - \frac{C^2}{A}}{G_\infty - \frac{\phi^2}{W}} \\ &= \frac{2\left[\left(\frac{1}{W} + \frac{\eta^2}{V}\right) - \frac{\phi^2}{W^2}\left(\frac{1}{W} + \frac{\eta^2}{V}\right)^{-1}\right]}{\left(\frac{1-\phi^2}{W} + \frac{\eta^2}{V}\right) + \sqrt{\left(\frac{1-\phi^2}{W} + \frac{\eta^2}{V}\right)^2 + 4\frac{\phi^2\eta^2}{VW}}} . \end{aligned}$$

The asymptotic precision ratio can be expressed as a function of V/W

$$\lim_{N\to\infty} \frac{\text{Var } \tilde{\beta}_{N|N}}{\text{Var } \tilde{\beta}^l_{N|N}} = \frac{2\left[\left(\frac{V}{W}+\eta^2\right) - \phi^2\left(\frac{V}{W}\right)^2\left(\frac{V}{W}+\eta^2\right)^{-1}\right]}{\left((1-\phi^2)\frac{V}{W}+\eta^2\right) + \sqrt{\left((1-\phi^2)\frac{V}{W}+\eta^2\right)^2 + 4\phi^2\eta^2\frac{V}{W}}} \ .$$

If $\phi^2/W \approx \eta^2/V$ then the asymptotic precision ratio is approximated by

$$\lim_{N\to\infty} \frac{\text{Var } \tilde{\beta}_{N|N}}{\text{Var } \tilde{\beta}^l_{N|N}} \approx \frac{2\left(1+\frac{\phi^4}{1+\phi^2}\right)}{1+\sqrt{1+4\phi^4}} \in (.927, 1] \ .$$

If $V/W = 0$ then the asymptotic precision ratio is 1. Figure 20 graphs a family of curves for asymptotic precision ratios where $|\phi| \in [0,1]$, $\eta = 1$, and where the curves correspond to $V/W = .5, 1, 3, 10, 50$ starting from the top right. It is interesting to note that the asymptotic precision ratio remains above .9 for $|\phi| \in [0, .8]$ in all curves. The next section generalizes the simple partial smoother estimate introduced in this section.

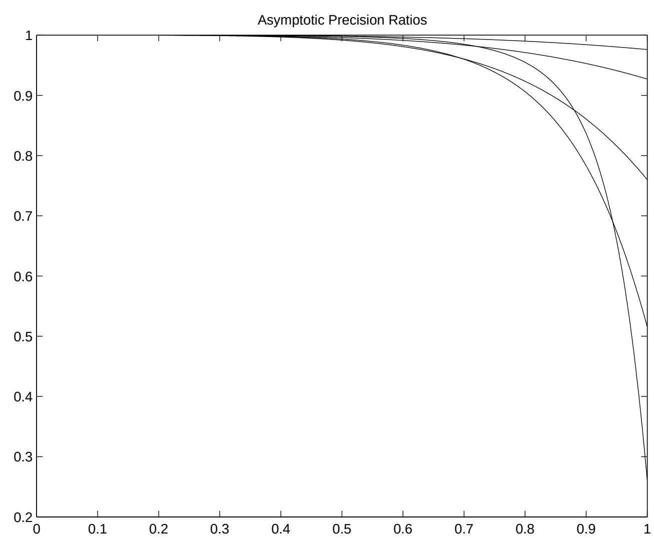

Figure 20: Asymptotic precision ratios of Kalman filter precisions versus partial smoother precisions where $|\phi| \in [0,1]$, $\eta = 1$, and where the different curves represent $V/W = .5, 1, 3, 10, 50$ starting from the top right.

3.5.2 A General Partial Smoother

As a general example of a partial smoother given the linear Gaussian state space model (91), consider a collection of new partial smoothers $\hat{\boldsymbol{\beta}}^m_{0:N} \equiv \{\hat{\beta}^m_{t|N} : t = 0, \ldots, N\}$ that satisfy the following new system of state estimating equations

$$\begin{aligned}
-\frac{1}{W}\left(\hat{\beta}^m_{N|N} - \phi\hat{\beta}^m_{N-1|N}\right) + \frac{\eta}{V}\left(Y_N - \eta\hat{\beta}^m_{N|N}\right) &= 0 \\
C_t\left(\hat{\beta}^m_{t+1|N} - \phi\hat{\beta}^m_{t|N}\right) - \frac{1}{W}\left(\hat{\beta}^m_{t|N} - \phi\hat{\beta}^m_{t-1|N}\right) + \frac{\eta}{V}\left(Y_t - \eta\hat{\beta}^m_{t|N}\right) &= 0 \\
t = N-1, \ldots, 1 & \\
C_0\left(\hat{\beta}^m_{1|N} - \phi\hat{\beta}^m_{0|N}\right) - \frac{1}{W_0}\left(\hat{\beta}^m_{0|N} - b_0\right) &= 0
\end{aligned}$$

where each $C_t \in \{0, \phi/W\}$ for $t = 0, \ldots, N-1$. This system of state estimating equations is written in matrix notation as

$$\begin{bmatrix}
A & -C & & & \\
-C_{N-1} & B_{N-1} & -C & & \\
 & & \ddots & & \\
 & & -C_1 & B_1 & -C \\
 & & & -C_0 & D_*
\end{bmatrix}
\begin{pmatrix}
\hat{\beta}^m_{N|N} \\ \hat{\beta}^m_{N-1|N} \\ \vdots \\ \hat{\beta}^m_{1|N} \\ \hat{\beta}^m_{0|N}
\end{pmatrix}
=
\begin{pmatrix}
\frac{\eta}{V}Y_N \\ \frac{\eta}{V}Y_{N-1} \\ \vdots \\ \frac{\eta}{V}Y_1 \\ \frac{b_0}{W_0}
\end{pmatrix}$$

$$A = \frac{1}{W} + \frac{\eta}{V},\ C = \frac{\phi}{W},\ D_0 = \frac{1}{W_0}$$

$$B_t = A + \phi C_t,\ t = N-1, \ldots, 1$$

$$D_* = D_0 + \phi C_0\ .$$

is represented in matrix symbology as

$$\boldsymbol{K}_m \hat{\boldsymbol{\beta}}^m_{N:0} = \boldsymbol{Y}^*_{N:0} \tag{113}$$

and is different from the system of state estimating equations associated with the complete state space smoother (92) when $C_t = 0$ for any $t \in \{0, \ldots, N-1\}$. The matrix $\boldsymbol{K}_m$ has the following partition for some $r \in \{1, \ldots, N\}$

$$\boldsymbol{K}_m = \begin{bmatrix}
\boldsymbol{M}^m_r & -\boldsymbol{C}^m_{r-1} & & & \\
 & \boldsymbol{M}^m_{r-1} & -\boldsymbol{C}^m_{r-2} & & \\
 & & \ddots & & \\
 & & & \boldsymbol{M}^m_2 & -\boldsymbol{C}^m_1 \\
 & & & & \boldsymbol{M}^m_1
\end{bmatrix}$$

where for $j = 1, \ldots, r$

$$
\begin{aligned}
\boldsymbol{M}_j^m &= \begin{bmatrix} A & -C & & & \\ -C & B & -C & & \\ & & \ddots & & \\ & & -C & B & -C \\ & & & -C & B_j^* \end{bmatrix} \in \mathbb{R}^{n_j \times n_j},\ j = 1, \ldots, r \\
\boldsymbol{C}_j^m &= \begin{bmatrix} 0 & & \ldots & 0 \\ \vdots & & & \vdots \\ 0 & & & \\ C & 0 & \ldots & 0 \end{bmatrix} \in \mathbb{R}^{n_{j+1} \times n_j},\ j = 1, \ldots, r-1 \\
B_j^* &= \begin{cases} B & : j \in \{r, \ldots, 2\} \\ D_* & : j = 1 \end{cases} .
\end{aligned}
$$

Each $\boldsymbol{M}_j^m$ is positive definite and invertible. Solving $\boldsymbol{K}_m \boldsymbol{X}_{r:1} = \boldsymbol{0}$ with $\boldsymbol{X}_{r:1} = (\boldsymbol{X}_j' \in \mathbb{R}^{1 \times n_j}, j = r, \ldots, 1)'$ results in the following r equations

$$
\begin{aligned}
\boldsymbol{M}_1^m \boldsymbol{X}_1 &= \boldsymbol{0} \\
\boldsymbol{M}_j^m \boldsymbol{X}_j - \boldsymbol{C}_{j-1}^m \boldsymbol{X}_{j-1} &= \boldsymbol{0},\ j = 2, \ldots, r .
\end{aligned}
$$

The previous display shows that $\boldsymbol{K}_m$ has full column rank since the only solution of $\boldsymbol{K}_m \boldsymbol{X}_{r:1} = \boldsymbol{0}$ is $\boldsymbol{X}_{r:1} = \boldsymbol{0}$. A similar analysis with respect to $\boldsymbol{K}_m' \boldsymbol{X}_{r:1} = \boldsymbol{0}$ starting with $\boldsymbol{M}_r^m \boldsymbol{X}_r = \boldsymbol{0}$ shows that $\boldsymbol{K}_m$ has full row rank. Hence $\boldsymbol{K}_m$ is invertible and the partial smoothers $\hat{\boldsymbol{\beta}}_{0:N}^m$ satisfy the following system of state estimating equations

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_1^m &= (\boldsymbol{M}_1^m)^{-1} \boldsymbol{Y}_1^m \\
\hat{\boldsymbol{\beta}}_j^m &= \left(\boldsymbol{M}_j^m\right)^{-1} \left(\boldsymbol{Y}_j^m + \boldsymbol{C}_{j-1}^m \hat{\boldsymbol{\beta}}_{j-1}^m\right),\ j = 2, \ldots, r
\end{aligned}
$$

where the vector of partial smoothers and the vector of observations are partitioned as follows

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{N:0}^m &= \left(\hat{\boldsymbol{\beta}}_j^{m\prime} \in \mathbb{R}^{1 \times n_j} : j = r, \ldots, 1\right)' \\
\boldsymbol{Y}_{N:0}^* &= \left(\boldsymbol{Y}_j^{m\prime} \in \mathbb{R}^{1 \times n_j} : j = r, \ldots, 1\right)' .
\end{aligned}
$$

It is easy to see that each of the partial smoothers $\hat{\beta}_t \in \hat{\boldsymbol{\beta}}_j^m$ depend on the observations $Y_t \in \{\boldsymbol{Y}_1^m, \ldots, \boldsymbol{Y}_j^m\}$, $j = 1, \ldots, r$.

Substituting the states $\boldsymbol{\beta}_{N:0} \equiv (\beta_N, \ldots, \beta_0)'$ for the partial smoothers $\hat{\boldsymbol{\beta}}^m_{N:0}$ in (113) and applying the linear Gaussian state space model (91) shows

$$\boldsymbol{K}_m \boldsymbol{\beta}_{N:0} - \boldsymbol{Y}^*_{N:0} = \begin{pmatrix} \frac{1}{W} w_N - \frac{\eta}{V} v_N \\ -C_{N-1} w_N + \frac{1}{W} w_{N-1} - \frac{\eta}{V} v_{N-1} \\ \vdots \\ -C_1 w_2 + \frac{1}{W} w_1 - \frac{\eta}{V} v_1 \\ -C_0 w_1 + \frac{1}{W_0} (\beta_0 - b_0) \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{T}_m\right)$$

where $\boldsymbol{T}_m$ is a tridiagonal covariance matrix

$$\boldsymbol{T}_m = \begin{bmatrix} A & -C_{N-1} & & & \\ -C_{N-1} & B_{N-1} & -C_{N-2} & & \\ & & \ddots & & \\ & & -C_1 & B_1 & -C_0 \\ & & & -C_0 & D_* \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}^m_r & & \\ & \ddots & \\ & & \boldsymbol{M}^m_1 \end{bmatrix} .$$

Define the associated collection of partial smoother residuals as $\tilde{\boldsymbol{\beta}}^m_{0:N} \equiv \{\tilde{\beta}^m_{t|N} \equiv \beta_t - \hat{\beta}^m_{t|N} : t = 0, \ldots, N\}$. Hence the partial smoother residuals $\tilde{\boldsymbol{\beta}}^m_{0:N}$ satisfy the following relationship

$$\begin{aligned} \boldsymbol{K}_m \tilde{\boldsymbol{\beta}}^m_{N:0} &\sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{T}_m\right) \text{ or } \tilde{\boldsymbol{\beta}}^m_{N:0} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{M}^{-1}_m\right) \\ \tilde{\boldsymbol{\beta}}^m_{N:0} &\equiv \left(\tilde{\beta}^m_{t|N} : t = N, \ldots, 0\right)' \\ \boldsymbol{M}^{-1}_m &= \left(\boldsymbol{K}_m\right)^{-1} \boldsymbol{T}_m \left(\boldsymbol{K}'_m\right)^{-1} . \end{aligned}$$

The following analysis shows that the precisions associated with the general partial smoothers $\tilde{\beta}^m_{t|N}$ are lower bounded by the precisions associated with the state space smoothers $\tilde{\beta}_{t|N}$ and are upper bounded by the precisions associated with the simple partial smoothers $\tilde{\beta}^l_{t|t}$

$$\mathrm{Var}\ \tilde{\beta}_{t|N} \leq \mathrm{Var}\ \tilde{\beta}^m_{t|N} \leq \mathrm{Var}\ \tilde{\beta}^l_{t|t},\ t = 0, \ldots, N .$$

Proposition 3.10. *Given the linear Gaussian state space model in (91) then*

$$\mathit{Var}\ \tilde{\boldsymbol{\beta}}_{N:0} \leq \mathit{Var}\ \tilde{\boldsymbol{\beta}}^m_{N:0}$$

where $\tilde{\boldsymbol{\beta}}_{N:0} = (\tilde{\beta}_{t|N} : t = N, \ldots, 0)'$ *is a vector of state space smoother residuals and where* $\tilde{\boldsymbol{\beta}}^m_{N:0} = (\tilde{\beta}^m_{t|N} : t = N, \ldots, 0)'$ *is a vector of partial smoother residuals.*

Proof: The result is proven by showing

$$\boldsymbol{M}_N = (\text{Var } \tilde{\boldsymbol{\beta}}_{N:0})^{-1} \geq \boldsymbol{M}_m = (\text{Var } \tilde{\boldsymbol{\beta}}^m_{N:0})^{-1} .$$

Let $\boldsymbol{K}_m = \boldsymbol{T}_m + \boldsymbol{\Delta}_m$ in order to show that

$$\begin{aligned}
\boldsymbol{M}_m &= \boldsymbol{K}'_m \boldsymbol{T}_m^{-1} \boldsymbol{K}_m = \left(\boldsymbol{T}_m + \boldsymbol{\Delta}'_m\right) \boldsymbol{T}_m^{-1} \left(\boldsymbol{T}_m + \boldsymbol{\Delta}_m\right) \\
&= \left(\boldsymbol{T}_m + \boldsymbol{\Delta}_m + \boldsymbol{\Delta}'_m\right) + \boldsymbol{\Delta}'_m \boldsymbol{T}_m^{-1} \boldsymbol{\Delta}_m \\
&\equiv \boldsymbol{M}^1_m + \boldsymbol{M}^2_m \\
\boldsymbol{\Delta}_m &\equiv \begin{bmatrix} \mathbf{0} & -\boldsymbol{C}^m_{r-1} & & \\ & \ddots & \ddots & \\ & & \mathbf{0} & -\boldsymbol{C}^m_1 \\ & & & \mathbf{0} \end{bmatrix}
\end{aligned}$$

where

$$\boldsymbol{M}^1_m = \begin{bmatrix} \boldsymbol{M}^m_r & -\boldsymbol{C}^m_{r-1} & & & \\ -\boldsymbol{C}^{m\prime}_{r-1} & \boldsymbol{M}^m_{r-1} & -\boldsymbol{C}^m_{r-2} & & \\ & & \ddots & & \\ & & -\boldsymbol{C}^{m\prime}_2 & \boldsymbol{M}^m_2 & -\boldsymbol{C}^m_1 \\ & & & -\boldsymbol{C}^{m\prime}_1 & \boldsymbol{M}^m_1 \end{bmatrix}$$

$$\boldsymbol{M}^2_m = \begin{bmatrix} \mathbf{0} & & & \\ & \boldsymbol{C}^{m\prime}_{r-1} \left(\boldsymbol{M}^m_r\right)^{-1} \boldsymbol{C}^m_{r-1} & & \\ & & \ddots & \\ & & & \boldsymbol{C}^{m\prime}_1 \left(\boldsymbol{M}^m_2\right)^{-1} \boldsymbol{C}^m_1 \end{bmatrix} .$$

It is easy to see that $\boldsymbol{M}^1_m$ has the following tridiagonal structure

$$\boldsymbol{M}^1_m = \begin{bmatrix} A & -C & & & \\ -C & B_{N-1} & -C & & \\ & & \ddots & & \\ & & -C & B_1 & -C \\ & & & -C & D_* \end{bmatrix} .$$

In order to analyze the structure of $\boldsymbol{M}^2_m$, let $(\boldsymbol{M}^m_j)^{-1} = [z^j_{i_1,i_2} : i_1, i_2 = 1, \ldots, n_j]$, $j = 2, \ldots, r$. Hence each of the non-zero diagonal submatrices of $\boldsymbol{M}^2_m$ have the following structure for $j = 2, \ldots, r$

$$\boldsymbol{C}^{m\prime}_{j-1} \left(\boldsymbol{M}^m_j\right)^{-1} \boldsymbol{C}^m_{j-1} = C^2 z^j_{n_j,n_j} \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & & \vdots \\ 0 & & \ldots & 0 \end{bmatrix} \in \mathbb{R}^{n_{j-1} \times n_{j-1}} .$$

Similar to the result in Lemma 3.2, Gaussian elimination of $\boldsymbol{M}_j^m \boldsymbol{Z}_j = \boldsymbol{e}_{n_j}$ where $\boldsymbol{Z}_j = (z_{i,n_j}^j : i = 1, \ldots, n_j)'$ shows

$$z_{n_j,n_j}^j = \frac{1}{G_{n_j}} \text{ for } j = 2, \ldots, r \ .$$

Hence the matrix $\boldsymbol{M}_m$ has the following structure

$$\boldsymbol{M}_m = \begin{bmatrix} \boldsymbol{M}_r^* & -\boldsymbol{C}_{r-1}^m & & & \\ -\boldsymbol{C}_{r-1}^{m\prime} & \boldsymbol{M}_{r-1}^* & -\boldsymbol{C}_{r-2}^m & & \\ & & \ddots & & \\ & & -\boldsymbol{C}_2^{m\prime} & \boldsymbol{M}_2^* & -\boldsymbol{C}_1^m \\ & & & -\boldsymbol{C}_1^{m\prime} & \boldsymbol{M}_1^* \end{bmatrix}$$

where each of the diagonal submatrices in $\boldsymbol{M}_m$ have a tridiagonal structure

$$\boldsymbol{M}_j^* = \begin{bmatrix} A + \frac{C^2}{G_{n_{j+1}}} & -C & & & \\ -C & B & -C & & \\ & & \ddots & & \\ & & -C & B & -C \\ & & & -C & B_j^* \end{bmatrix} \in \mathbb{R}^{n_j \times n_j}, \ j = 1, \ldots, r-1$$
$$\boldsymbol{M}_r^* = \boldsymbol{M}_r^m \ .$$

Equation (101) from Lemma 3.5 showed $A \leq G_j < B$ for $j \in [1, \ldots, \infty)$. This earlier result implies

$$A + \frac{C^2}{G_j} \leq A + \frac{C^2}{G_1} < B \text{ for } j = 1, \ldots$$

which in turn is used to show $\boldsymbol{M}_N \geq \boldsymbol{M}_m$. Hence the result is proven since $\boldsymbol{M}_N \geq \boldsymbol{M}_m$ implies $\boldsymbol{M}_N^{-1} \leq \boldsymbol{M}_m^{-1}$. ■

Proposition 3.11. *Given the linear Gaussian state space model in (91) then*

$$Var\ \tilde{\beta}_{t|N}^m \leq Var\ \tilde{\beta}_{t|t}^l \text{ for } t = 0, \ldots, N$$

where $\tilde{\boldsymbol{\beta}}_{0:N}^m \equiv \{\tilde{\beta}_{t|N}^m : t = 0, \ldots, N\}$ *is a collection of the general partial smoother residuals and where* $\tilde{\boldsymbol{\beta}}_{0:N}^l \equiv \{\tilde{\beta}_{t|t}^l : t = 0, \ldots, N\}$ *is a collection of the simple partial smoother residuals.*

Proof: Partition the general partial smoother residuals into

$$\tilde{\boldsymbol{\beta}}^m_{N:0} = \left(\tilde{\boldsymbol{\beta}}^{m\prime}_j \equiv \left(\tilde{\beta}^m_{j,n_j}, \ldots, \tilde{\beta}^m_{j,1}\right) : j = r, \ldots, 1\right)'$$

such that the distribution for $\boldsymbol{K}_m \tilde{\boldsymbol{\beta}}^m_{N:0}$ satisfies

$$\begin{gathered}
\boldsymbol{M}^m_1 \tilde{\boldsymbol{\beta}}^m_1 \equiv \boldsymbol{W}^m_1 \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{M}^m_1\right) \\
\boldsymbol{M}^m_j \tilde{\boldsymbol{\beta}}^m_j - \boldsymbol{C}^m_{j-1} \tilde{\boldsymbol{\beta}}^m_{j-1} \equiv \boldsymbol{W}^m_j \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{M}^m_j\right),\ j = 2, \ldots, r
\end{gathered}$$

where the partition of general partial smoother residuals satisfy

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}^m_1 &= \left(\boldsymbol{M}^m_1\right)^{-1} \boldsymbol{W}^m_1 \\
\tilde{\boldsymbol{\beta}}^m_j &= \left(\boldsymbol{M}^m_j\right)^{-1} \boldsymbol{W}^m_j + \left(\boldsymbol{M}^m_j\right)^{-1} \boldsymbol{C}^m_{j-1} \tilde{\boldsymbol{\beta}}^m_{j-1},\ j = 2, \ldots, r
\end{aligned}$$

and where the random vector sequence $\{\boldsymbol{W}^m_j,\ j = 1, \ldots, r\}$ is independent.

In a similar manner, partition the simple partial smoother residuals into

$$\tilde{\boldsymbol{\beta}}^l_{N:0} = \left(\tilde{\boldsymbol{\beta}}^{l\prime}_j \equiv \left(\tilde{\beta}^l_{j,n_j}, \ldots, \tilde{\beta}^l_{j,1}\right) : j = r, \ldots, 1\right)'$$

and partition the coefficient matrix $\boldsymbol{U}_l$ and the covariance matrix $\boldsymbol{D}_l$ into

$$\begin{aligned}
\boldsymbol{U}_l &= \begin{bmatrix} \boldsymbol{U}^l_r & -\boldsymbol{C}^m_{r-1} & & \\ & \ddots & & \\ & & \boldsymbol{U}^l_2 & -\boldsymbol{C}^m_1 \\ & & & \boldsymbol{U}^l_1 \end{bmatrix},\ \boldsymbol{U}^l_j \in \mathbb{R}^{n_j \times n_j},\ j = 1, \ldots, r \\
\boldsymbol{D}_l &= \begin{bmatrix} \boldsymbol{D}^l_r & & \\ & \ddots & \\ & & \boldsymbol{D}^l_1 \end{bmatrix},\ \boldsymbol{D}^l_j \in \mathbb{R}^{n_j \times n_j},\ j = 1, \ldots, r\ .
\end{aligned}$$

such that the distribution for $\boldsymbol{U}_l \tilde{\boldsymbol{\beta}}^l_{N:0}$ satisfies

$$\begin{gathered}
\boldsymbol{U}^l_1 \tilde{\boldsymbol{\beta}}^l_1 \equiv \boldsymbol{W}^l_1 \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{D}^l_1\right) \\
\boldsymbol{U}^l_j \tilde{\boldsymbol{\beta}}^l_j - \boldsymbol{C}^m_{j-1} \tilde{\boldsymbol{\beta}}^l_{j-1} \equiv \boldsymbol{W}^l_j \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{D}^l_j\right),\ j = 2, \ldots, r
\end{gathered}$$

where the partition of simple partial smoother residuals satisfy

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}^l_1 &= \left(\boldsymbol{U}^l_1\right)^{-1} \boldsymbol{W}^l_1 \\
\tilde{\boldsymbol{\beta}}^l_j &= \left(\boldsymbol{U}^l_j\right)^{-1} \boldsymbol{W}^l_j + \left(\boldsymbol{U}^l_j\right)^{-1} \boldsymbol{C}^m_{j-1} \tilde{\boldsymbol{\beta}}^l_{j-1},\ j = 2, \ldots, r
\end{aligned}$$

and where the random vector sequence $\{\boldsymbol{W}_j^l,\, j=1,\ldots,r\}$ is independent.

Let $(\boldsymbol{M}_j^m)^{-1}=[z_{i_1,i_2}^j : i_1,i_2=1,\ldots,n_j]$ for $j=1,\ldots,r$ such that

$$\left(\boldsymbol{M}_j^m\right)^{-1}\boldsymbol{C}_{j-1}^m\tilde{\boldsymbol{\beta}}_{j-1}^m = C\tilde{\beta}_{j-1,n_{j-1}}^m\boldsymbol{Z}_j \tag{114}$$

$$\boldsymbol{Z}_j=\left(z_{i,n_j}^j,\, i=1,\ldots,n_j\right)'$$

and let $(\boldsymbol{U}_j^l)^{-1}=[q_{i_1,i_2}^j : i_1,i_2=1,\ldots,n_j]$ for $j=1,\ldots,r$ such that

$$\left(\boldsymbol{U}_j^l\right)^{-1}\boldsymbol{C}_{j-1}^m\tilde{\boldsymbol{\beta}}_{j-1}^l = C\tilde{\beta}_{j-1,n_{j-1}}^l\boldsymbol{Q}_j \tag{115}$$

$$\boldsymbol{Q}_j=\left(q_{i,n_j}^j,\, i=1,\ldots,n_j\right)' \; .$$

The two covariance matrices for (114) and (115) follow directly for $j=2,\ldots,r$

$$\mathrm{Var}\left[\left(\boldsymbol{M}_j^m\right)^{-1}\boldsymbol{C}_{j-1}^m\tilde{\boldsymbol{\beta}}_{j-1}^m\right]=C^2\mathrm{Var}\left(\tilde{\beta}_{j-1,n_{j-1}}^m\right)\boldsymbol{Z}_j\boldsymbol{Z}_j'$$

$$\mathrm{Var}\left[\left(\boldsymbol{U}_j^l\right)^{-1}\boldsymbol{C}_{j-1}^m\tilde{\boldsymbol{\beta}}_{j-1}^l\right]=C^2\mathrm{Var}\left(\tilde{\beta}_{j-1,n_{j-1}}^l\right)\boldsymbol{Q}_j\boldsymbol{Q}_j' \; .$$

Hence the covariance matrices of $\tilde{\boldsymbol{\beta}}_j^m$ for $j=1,\ldots,r$ are

$$\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_1^m\right)=\left(\boldsymbol{M}_1^m\right)^{-1} \tag{116}$$

$$\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_j^m\right)=\left(\boldsymbol{M}_j^m\right)^{-1}+C^2\mathrm{Var}\left(\tilde{\beta}_{j-1,n_{j-1}}^m\right)\boldsymbol{Z}_j\boldsymbol{Z}_j' \tag{117}$$

and the covariance matrices of $\tilde{\boldsymbol{\beta}}_j^l$ for $j=1,\ldots,r$ are

$$\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_1^l\right)=\left(\boldsymbol{U}_1^l\right)^{-1}\boldsymbol{D}_1^l\left(\boldsymbol{U}_1^{l\prime}\right)^{-1} \tag{118}$$

$$\mathrm{Var}\left(\tilde{\boldsymbol{\beta}}_j^l\right)=\left(\boldsymbol{U}_j^l\right)^{-1}\boldsymbol{D}_j^l\left(\boldsymbol{U}_j^{l\prime}\right)^{-1}+C^2\mathrm{Var}\left(\tilde{\beta}_{j-1,n_{j-1}}^l\right)\boldsymbol{Q}_j\boldsymbol{Q}_j' \; . \tag{119}$$

Proposition 3.10 with $\tilde{\boldsymbol{\beta}}_{N:0}=\tilde{\boldsymbol{\beta}}_j^m$ and $\tilde{\boldsymbol{\beta}}_{N:0}^m=\tilde{\boldsymbol{\beta}}_j^l$ shows for $j=1,\ldots,r$

$$\left(\boldsymbol{M}_j^m\right)^{-1}\le\left(\boldsymbol{U}_j^l\right)^{-1}\boldsymbol{A}_j^l\left(\boldsymbol{U}_j^{l\prime}\right)^{-1} \; . \tag{120}$$

In view of the three previous displays, equations (116) through (120), the result is proven if the following diagonal covariance inequality is true for $j=2,\ldots,r$ and $i=1,\ldots,n_j$

$$C^2\mathrm{Var}\left(\tilde{\beta}_{j-1,n_{j-1}}^m\right)\left(z_{i,i}^j\right)^2\le C^2\mathrm{Var}\left(\tilde{\beta}_{j-1,n_{j-1}}^l\right)\left(q_{i,i}^j\right)^2 \; . \tag{121}$$

Similar to the result in Lemma 3.2, Gaussian elimination of $\boldsymbol{M}_j^m \boldsymbol{Z}_j = \boldsymbol{e}_{n_j}$ where $\boldsymbol{Z}_j = (z_{i,n_j}^j : i = 1, \ldots, n_j)'$ shows for $j = 2, \ldots, r$

$$\begin{aligned}\boldsymbol{Z}_j &= \left(\frac{C}{G_1} \times \cdots \times \frac{C}{G_{n_j-1}} \times \frac{1}{G_{n_j}}, \ldots, \frac{C}{G_{n_j-1}} \times \frac{1}{G_{n_j}}, \frac{1}{G_{n_j}} \right)' \\ &= \frac{1}{G_{n_j}} \left(\frac{C}{G_1} \times \cdots \times \frac{C}{G_{n_j-1}}, \ldots, \frac{C}{G_{n_j-1}}, 1 \right)' .\end{aligned}$$

The following display gives the structure of the coefficient matrix $\boldsymbol{U}_j^l$ and its inverse for $j = 2, \ldots, r$

$$\boldsymbol{U}_j^l = \begin{bmatrix} A & -C & & \\ & \ddots & & \\ & & A & -C \\ & & & A \end{bmatrix}, \left(\boldsymbol{U}_j^l\right)^{-1} = \begin{bmatrix} \frac{1}{A} & \frac{C}{A^2} & \frac{C^2}{A^3} & \cdots & & & \\ & \frac{1}{A} & \frac{C}{A^2} & \frac{C^2}{A^3} & \cdots & & \\ & & & \ddots & & & \\ & & & & \frac{1}{A} & \frac{C}{A^2} & \frac{C^2}{A^3} \\ & & & & & \frac{1}{A} & \frac{C}{A^2} \\ & & & & & & \frac{1}{A} \end{bmatrix}$$

which shows for $j = 2, \ldots, r$

$$\boldsymbol{Q}_j = \frac{1}{A} \left(\left(\frac{C}{A} \right)^{n_j - 1}, \ldots, \left(\frac{C}{A} \right)^0 \right)' .$$

Hence the following diagonal inequality is true since $A \leq G_k$ for $k = 1, \ldots, N$

$$\left(z_{i,i}^j \right)^2 \leq \left(q_{i,i}^j \right)^2 \text{ for } j = 1, \ldots, r \text{ and } i = 1, \ldots, n_j . \tag{122}$$

The covariance inequality (120) for $j = 1$ together with the initial partial smoothers equations (116) and (118) shows Var $\tilde{\boldsymbol{\beta}}_1^m \leq$ Var $\tilde{\boldsymbol{\beta}}_1^l$, which proves the result that Var $\tilde{\beta}_{1,i}^m \leq$ Var $\tilde{\beta}_{1,i}^l$ for $i = 1, \ldots, n_1$. This inequality together with the diagonal inequality (122) shows the diagonal covariance inequality (121) for $j = 2$. For $j = 2$, the combination of inequalities (120) and (121) together with the partial smoothers equations (117) and (119) proves the result that Var $\tilde{\beta}_{2,i}^m \leq$ Var $\tilde{\beta}_{2,i}^l$ for $i = 1, \ldots, n_2$. Induction is used to show the diagonal covariance inequality (121) for $j = 3, \ldots, r$. For $j = 3, \ldots, r$, the combination of inequalities (120) and (121) proves the result that Var $\tilde{\beta}_{j,i}^m \leq$ Var $\tilde{\beta}_{j,i}^l$ for $i = 1, \ldots, n_j$. Hence the complete result has been proven. ■

3.5.3 A Partial Smoother With Constant Partition Size

As a special case of the general partial smoothers, let $\hat{\boldsymbol{\beta}}^{m_n}_{0:N} \equiv \{\hat{\beta}^{m_n}_0, \ldots, \hat{\beta}^{m_n}_N\}$ represent the collection of partial smoothers where each of the r partitions have the same size n such that $n_1 = n_2 = \cdots = n_r \equiv n > 1$. The general partial smoothers $\hat{\boldsymbol{\beta}}^{m_n}_{0:N}$, for the case where the partition size $n = 1$, are equivalent to the simple partial smoothers $\hat{\boldsymbol{\beta}}^{l}_{0:N}$. The partial smoothers $\hat{\boldsymbol{\beta}}^{m_n}_{0:N}$ satisfy the following system of state estimating equations

$$
\begin{aligned}
\boldsymbol{K}_{m_n}\hat{\boldsymbol{\beta}}^{m_n}_{N:0} &= \boldsymbol{Y}^*_{N:0} \\
\hat{\boldsymbol{\beta}}^{m_n}_{N:0} &\equiv \left(\hat{\boldsymbol{\beta}}^{m_n\prime}_j \equiv \left(\hat{\beta}_{j,n}, \ldots, \hat{\beta}_{j,1}\right) : j = r, \ldots, 1\right)' \equiv \left(\hat{\beta}^{m_n}_N, \ldots, \hat{\beta}^{m_n}_0\right)' \\
\boldsymbol{K}_{m_n} &\equiv \begin{bmatrix} \boldsymbol{M}^{m_n}_r & -\boldsymbol{C}^{m_n}_{r-1} & & \\ & \ddots & & \\ & & \boldsymbol{M}^{m_n}_2 & -\boldsymbol{C}^{m_n}_1 \\ & & & \boldsymbol{M}^{m_n}_1 \end{bmatrix} \\
\boldsymbol{M}^{m_n}_j &\equiv \boldsymbol{M}^{m}_j \in \mathbb{R}^{n\times n},\ j = 1, \ldots, r \\
\boldsymbol{C}^{m_n}_j &\equiv \boldsymbol{C}^{m}_j \in \mathbb{R}^{n\times n},\ j = 1, \ldots, r \\
\boldsymbol{M}^{m_n}_2 &= \cdots = \boldsymbol{M}^{m_n}_r\ .
\end{aligned}
$$

Let $\tilde{\boldsymbol{\beta}}^{m_n}_{0:N} \equiv \{\tilde{\beta}^{m_n}_0, \ldots, \tilde{\beta}^{m_n}_N\}$ represent the associated collection of partial smoother residuals that satisfy the following relationship

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}}^{m_n}_{N:0} &\sim \mathrm{N}\left(\boldsymbol{0}, (\boldsymbol{M}_{m_n})^{-1}\right) \\
\tilde{\boldsymbol{\beta}}^{m_n}_{N:0} &\equiv \left(\tilde{\boldsymbol{\beta}}^{m_n\prime}_j \equiv \left(\tilde{\beta}_{j,n}, \ldots, \tilde{\beta}_{j,1}\right) : j = r, \ldots, 1\right)' \equiv \left(\tilde{\beta}^{m_n}_N, \ldots, \tilde{\beta}^{m_n}_0\right)' \\
\boldsymbol{M}_{m_n} &\equiv \begin{bmatrix} \boldsymbol{M}^*_r & -\boldsymbol{C}^{m_n}_{r-1} & & & \\ -\boldsymbol{C}^{m_n\prime}_{r-1} & \boldsymbol{M}^*_{r-1} & -\boldsymbol{C}^{m_n}_{r-2} & & \\ & & \ddots & & \\ & & -\boldsymbol{C}^{m_n\prime}_2 & \boldsymbol{M}^*_2 & -\boldsymbol{C}^{m_n}_1 \\ & & & -\boldsymbol{C}^{m_n\prime}_1 & \boldsymbol{M}^*_1 \end{bmatrix} \\
\boldsymbol{M}^*_j &\in \mathbb{R}^{n\times n},\ j = 1, \ldots, r \\
\boldsymbol{M}^*_2 &= \cdots = \boldsymbol{M}^*_r\ .
\end{aligned}
$$

As a special case of the general partial smoothers, the results of Propositions 3.10 and Lemma 3.11 are valid with respect to the partial smoother residuals $\tilde{\boldsymbol{\beta}}^{m_n}_{0:N}$. Due to the constant partition size n, it is possible to find the asymptotic precision for the partitioned partial smoothers $\tilde{\boldsymbol{\beta}}^{m_n}_r$.

Theorem 3.2. *Given the linear Gaussian state space model (91) with $N = rn - 1$, then the precision of the constant partitioned partial smoothers $\hat{\boldsymbol{\beta}}_r^{m_n}$ converges to a finite covariance matrix as $r \to \infty$*

$$\begin{aligned}
Var\ \tilde{\boldsymbol{\beta}}_r^{m_n} &\to \boldsymbol{P}_*^{m_n} = (\boldsymbol{M}_2^{m_n})^{-1} + C^2 P_{n,n}^{m_n} \boldsymbol{Z}_n \boldsymbol{Z}_n' \\
Var\ \tilde{\beta}_{r,n}^{m_n} &\to P_{n,n}^{m_n} \equiv \frac{z_{1,1}}{1 - C^2 z_{1,n}^2} \\
(\boldsymbol{M}_2^{m_n})^{-1} &\equiv [z_{i_1,i_2} : i_1, i_2 = 1, \dots n] \\
\boldsymbol{Z}_n &\equiv (z_{i,n} : i = 1, \dots, n)'
\end{aligned}$$

where

$$\begin{aligned}
z_{1,1} &= \left(A - \frac{C^2}{G_{n-1}^*(B)} \right)^{-1} \\
\boldsymbol{Z}_n &= \frac{1}{G_n} \left(\frac{C}{G_1} \times \cdots \times \frac{C}{G_{n-1}}, \dots, \frac{C}{G_{n-1}}, 1 \right)' \\
G_j^*(D) &= \begin{cases} D & : j = 1 \\ B - \frac{C^2}{G_{j-1}^*} & : j > 1 \end{cases} .
\end{aligned}$$

Proof: Equation (117) within Proposition 3.11 gives the precision for $\hat{\boldsymbol{\beta}}_r^{m_n}$ as

$$\text{Var}\left(\tilde{\boldsymbol{\beta}}_r^{m_n}\right) = (\boldsymbol{M}_2^{m_n})^{-1} + C^2 \text{Var}\left(\tilde{\beta}_{r-1,n}^{m_n}\right) \boldsymbol{Z}_n \boldsymbol{Z}_n' \tag{123}$$

which shows that the precision for $\hat{\boldsymbol{\beta}}_{r,n}^{m_n}$, $r > 2$, is

$$\begin{aligned}
\text{Var}\left(\tilde{\beta}_{r,n}^{m_n}\right) &= z_{1,1} + C^2 z_{1,n}^2 \text{Var}\left(\tilde{\beta}_{r-1,n}^{m_n}\right) \\
&= z_{1,1}\left(1 + \left(C^2 z_{1,n}^2\right) + \left(C^2 z_{1,n}^2\right)^2 + \cdots + \left(C^2 z_{1,n}^2\right)^{r-3}\right) \\
&\quad + \left(C^2 z_{1,n}^2\right)^{r-2} \left(z_{1,1}^{(1)} + \left(C^2 z_{1,n}^2\right) \text{Var}\left(\tilde{\beta}_{1,n}^{m_n}\right)\right) \\
(\boldsymbol{M}_1^{m_n})^{-1} &\equiv \left[z_{i_1,i_2}^{(1)} : i_1, i_2 = 1, \dots n\right] .
\end{aligned}$$

The formula for $z_{1,1}$ is found by using Gaussian elimination to solve $\boldsymbol{M}_2^{m_n} \boldsymbol{Z}_1 = \boldsymbol{e}_1$ with $\boldsymbol{Z}_1 = (z_{1,1}, \dots, z_{n,1})'$. The formula for $\boldsymbol{Z}_n$ is found by using Gaussian elimination to solve $\boldsymbol{M}_2^{m_n} \boldsymbol{Z}_n = \boldsymbol{e}_n$. The formula for $z_{1,1}^{(1)}$ is found by using Gaussian elimination to solve $\boldsymbol{M}_1^{m_n} \boldsymbol{Z}_1^{(1)} = \boldsymbol{e}_1$ with $\boldsymbol{Z}_1^{(1)} = (z_{1,1}^{(1)}, \dots, z_{n,1}^{(1)})'$

$$z_{1,1}^{(1)} = \left(A - \frac{C^2}{G_{n-1}^*(D)} \right)^{-1} .$$

Hence the precision for $\hat{\boldsymbol{\beta}}_{r,n}^{m_n}$ converges to a finite limit as $r \to \infty$ since $|C^2 z_{1,n}^2| < 1$

$$\mathrm{Var}\left(\tilde{\beta}_{r,n}^{m_n}\right) \to \frac{z_{1,1}}{1 - C^2 z_{1,n}^2} \equiv P_{n,n}^{m_n} = z_{1,1} + C^2 z_{1,n}^2 P_{n,n}^{m_n} \ .$$

The result is proven using (123). ∎

3.5.4 Another Partial Smoother

In this section another partitioned sequence of partial smoothers $\hat{\boldsymbol{\beta}}_{0:N}^{s}$ is described that satisfies a system of state estimating equations different from the previous partial smoothers $\hat{\boldsymbol{\beta}}_{0:N}^{m}$ and different from the complete state space smoothers $\boldsymbol{\beta}_{0:N|N}^{k}$. The precisions associated with these partial smoothers $\hat{\boldsymbol{\beta}}_{0:N}^{s}$ are smaller than the precisions associated with a comparable partition of the previous partial smoothers $\hat{\boldsymbol{\beta}}_{0:N}^{m}$ and are larger than the precisions associated with the complete state space smoothers $\boldsymbol{\beta}_{0:N|N}^{k}$.

Using a constant partition size of $n+1$, divide the sequence of states $\boldsymbol{\beta}_{0:N} \equiv \{\beta_0, \ldots, \beta_N\}$ into r overlapping partitions as follows

$$\begin{aligned} \boldsymbol{\beta}_{r:1}^{s} &\equiv \left(\boldsymbol{\beta}_{j}^{s\prime} : j = r, \ldots, 1\right)' \\ \boldsymbol{\beta}_{j}^{s} &\equiv (\beta_{j,i} : i = n, \ldots, 0)' \equiv \left(\beta_{jn}, \ldots, \beta_{(j-1)n}\right)' \ . \end{aligned}$$

where each state partition $\{\boldsymbol{\beta}_{j}^{s} : 1 < j \leq r\}$ contains an initial state $\beta_{j,0}$ that corresponds to the last state $\beta_{j-1,n}$ from the previous partition

$$\beta_{j,0} \equiv \beta_{j-1,n} \equiv \beta_{(j-1)n}, \ j = 2, \ldots, r \ .$$

Also using a constant partition size of n, divide the sequence of observations $\boldsymbol{\mathcal{F}}_N \equiv \{Y_1, \ldots, Y_N\}$ into r non-overlapping partitions as follows

$$\begin{aligned} \boldsymbol{Y}_{N:1} &\equiv \left(\boldsymbol{Y}_{j}^{s\prime} \equiv (Y_{j,i} : i = n, \ldots, 1) : j = r, \ldots, 1\right)' \\ &\equiv (Y_N, \ldots, Y_1)' \ . \end{aligned}$$

Denote the partial smoothers as $\hat{\boldsymbol{\beta}}_{0:N}^{s} \equiv \{\hat{\beta}_0^s, \ldots, \hat{\beta}_N^s\}$ and the partial smoother residuals as $\tilde{\boldsymbol{\beta}}_{0:N}^{s} \equiv \{\tilde{\beta}_t^s \equiv \beta_t - \hat{\beta}_t^s : t = 0, \ldots, N\}$. Divide the partial smoothers and the partial smoother residuals using a constant partition size of $n+1$ such that each partition has an initial random variable

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^s_{r:1} &\equiv \left(\hat{\boldsymbol{\beta}}^{s\prime}_j \equiv \left(\hat{\beta}^s_{j,i} : i = n, \ldots, 0\right) : j = r, \ldots, 1\right)' \\
&\equiv \left(\left(\hat{\beta}^s_{jn}, \ldots, \hat{\beta}^s_{jn-n+1}, \hat{\beta}^s_{j,0}\right) : j = r, \ldots, 1\right)' \\
\hat{\beta}^s_0 &\equiv b_0 \\
\tilde{\boldsymbol{\beta}}^s_{r:1} &\equiv \left(\tilde{\boldsymbol{\beta}}^{s\prime}_j \equiv \left(\tilde{\beta}^s_{j,i} : i = n, \ldots, 0\right) : j = r, \ldots, 1\right)' \\
&\equiv \left(\left(\tilde{\beta}^s_{jn}, \ldots, \tilde{\beta}^s_{jn-n+1}, \tilde{\beta}^s_{j,0}\right) : j = r, \ldots, 1\right)' \\
\tilde{\beta}^s_0 &\equiv \beta_0 - \hat{\beta}^s_0
\end{aligned}$$

where the initial state of each smoother partition $\{\hat{\beta}^s_{j,0} : j = 1, \ldots, r\}$, will be used to estimate the corresponding initial state of each state partition $\{\beta_{j,0} : j = 1, \ldots, r\}$. Using the linear Gaussian state space model (91) shows that the first partition of states $\boldsymbol{\beta}^s_1 = (\beta_{1,n}, \ldots, \beta_{1,0})'$ satisfies the following system of equations

$$\begin{aligned}
-\frac{1}{W}\left(\beta_{1,n} - \phi\beta_{1,n-1}\right) + \frac{\eta}{V}\left(Y_{1,n} - \eta\beta_{1,n}\right)& \\
&= -\frac{1}{W}w_n + \frac{\eta}{V}v_n \\
\frac{\phi}{W}\left(\beta_{1,t+1} - \phi\beta_{1,t}\right) - \frac{1}{W}\left(\beta_{1,t} - \phi\beta_{1,t-1}\right) + \frac{\eta}{V}\left(Y_{1,t} - \eta\beta_{1,t}\right)& \\
&= \frac{\phi}{W}w_{t+1} - \frac{1}{W}w_t + \frac{\eta}{V}v_t \\
&\text{for } t = n-1, \ldots, 1 \\
\frac{\phi}{W}\left(\beta_{1,1} - \phi\beta_{1,0}\right) - \frac{1}{W_0}\left(\beta_{1,0} - b_0\right)& \\
&= \frac{\phi}{W}w_1 - \frac{1}{W_0}w_0 \, .
\end{aligned}$$

Let the first partition of partial smoothers $\hat{\boldsymbol{\beta}}^s_1 = (\hat{\beta}^s_{1,n}, \ldots, \hat{\beta}^s_{1,0})'$ satisfy the following system of state estimating equations

$$\begin{aligned}
-\frac{1}{W}\left(\hat{\beta}^s_{1,n} - \phi\hat{\beta}^s_{1,n-1}\right) + \frac{\eta}{V}\left(Y_{1,n} - \eta\hat{\beta}^s_{1,n}\right) &= 0 \\
\frac{\phi}{W}\left(\hat{\beta}^s_{1,t+1} - \phi\hat{\beta}^s_{1,t}\right) - \frac{1}{W}\left(\hat{\beta}^s_{1,t} - \phi\hat{\beta}^s_{1,t-1}\right) + \frac{\eta}{V}\left(Y_{1,t} - \eta\hat{\beta}^s_{1,t}\right) &= 0 \\
\text{for } t = n-1, \ldots, 1& \\
\frac{\phi}{W}\left(\hat{\beta}^s_{1,1} - \phi\hat{\beta}^s_{1,0}\right) - \frac{1}{W_0}\left(\hat{\beta}^s_{1,0} - b_0\right) &= 0 \, .
\end{aligned}$$

Each partial smoother $\hat{\beta}^s_{j,i} \in \hat{\boldsymbol{\beta}}^s_1$ depends on the observations $Y_t \in \boldsymbol{Y}^s_1$. The first partition of partial smoother residuals $\tilde{\boldsymbol{\beta}}^s_1 = (\tilde{\beta}^s_{1,n}, \ldots, \tilde{\beta}^s_{1,0})'$ has the following distribution

$$\boldsymbol{M}^s_1 \tilde{\boldsymbol{\beta}}^s_1 \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{M}^s_1\right) \text{ or } \tilde{\boldsymbol{\beta}}^s_1 \sim \mathrm{N}\left(\boldsymbol{0}, \left(\boldsymbol{M}^s_1\right)^{-1}\right)$$
$$\boldsymbol{M}^s_1 \equiv \boldsymbol{M}_n \in \mathbb{R}^{n+1 \times n+1}$$

where the precision for $\hat{\beta}^s_{1,n}$, as shown in Lemma 3.12, is

$$\begin{aligned}
\mathrm{Var}\ \tilde{\beta}^s_{1,n} &= \left(A - \frac{C^2}{G^*_n\left(D\right)}\right)^{-1} \\
G^*_k\left(D\right) &\equiv \begin{cases} D & : k = 1 \\ B - \frac{C^2}{G^*_{k-1}(D)} & : k > 1 \end{cases} .
\end{aligned}$$

The linear Gaussian state space model shows that each subsequent partition of states $\{\boldsymbol{\beta}^s_j = (\beta_{j,n}, \ldots, \beta_{j,0})' : j = 2, \ldots, r\}$ satisfies the following system of equations

$$\begin{aligned}
-\frac{1}{W}&\left(\beta_{j,n} - \phi\beta_{j,n-1}\right) + \frac{\eta}{V}\left(Y_{j,n} - \eta\beta_{j,n}\right) \\
&= -\frac{1}{W} w_{jn} + \frac{\eta}{V} v_{jn} \\
\frac{\phi}{W}&\left(\beta_{j,t+1} - \phi\beta_{j,t}\right) - \frac{1}{W}\left(\beta_{j,t} - \phi\beta_{j,t-1}\right) + \frac{\eta}{V}\left(Y_{j,t} - \eta\beta_{j,t}\right) \\
&= \frac{\phi}{W} w_{(j-1)n+t+1} - \frac{1}{W} w_{(j-1)n+t} + \frac{\eta}{V} v_{(j-1)n+t} \\
&\text{for } t = n-1, \ldots, 1 \\
\frac{\phi}{W}&\left(\beta_{j,1} - \phi\beta_{j,0}\right) - \left(A - \frac{C^2}{G^*_{(j-1)n}\left(D\right)}\right)\left(\beta_{j,0} - \beta_{j-1,n}\right) \\
&= \frac{\phi}{W} w_{(j-1)n+1} .
\end{aligned}$$

Let each subsequent partition of partial smoothers $\{\hat{\boldsymbol{\beta}}^s_j : j = 2, \ldots, r\}$ satisfy the following system of state estimating equations

$$
\begin{aligned}
-\frac{1}{W}\left(\hat{\beta}^s_{j,n}-\phi\hat{\beta}^s_{j,n-1}\right)+\frac{\eta}{V}\left(Y_{j,n}-\eta\hat{\beta}^s_{j,n}\right)&=0\\
\frac{\phi}{W}\left(\hat{\beta}^s_{j,t+1}-\phi\hat{\beta}^s_{j,t}\right)-\frac{1}{W}\left(\hat{\beta}^s_{j,t}-\phi\hat{\beta}^s_{j,t-1}\right)+\frac{\eta}{V}\left(Y_{j,t}-\eta\hat{\beta}^s_{j,t}\right)&=0\\
\text{for } t=n-1,\dots,1\quad&\\
\frac{\phi}{W}\left(\hat{\beta}^s_{j,1}-\phi\hat{\beta}^s_{j,0}\right)-\left(A-\frac{C^2}{G^*_{(j-1)n}(D)}\right)\left(\hat{\beta}^s_{j,0}-\hat{\beta}^s_{j-1,n}\right)&=0\;.
\end{aligned}
$$

It is easy to see that each partial smoother $\hat{\beta}^s_{j,i}\in\hat{\boldsymbol{\beta}}^s_j$ depends on the observations $Y_t\in\{\boldsymbol{Y}^s_1,\dots,\boldsymbol{Y}^s_j\}$, $j=2,\dots,r$. Each subsequent partition of partial smoother residuals $\{\tilde{\boldsymbol{\beta}}^s_j : j=2,\dots,r\}$ satisfies the following system of equations

$$
\begin{aligned}
&-\frac{1}{W}\left(\tilde{\beta}_{j,n}-\phi\tilde{\beta}_{j,n-1}\right)-\frac{\eta^2}{V}\tilde{\beta}_{j,n}\\
&\qquad\qquad=-\frac{1}{W}w_{jn}+\frac{\eta}{V}v_{jn}\\
&\frac{\phi}{W}\left(\tilde{\beta}_{j,t+1}-\phi\tilde{\beta}_{j,t}\right)-\frac{1}{W}\left(\tilde{\beta}_{j,t}-\phi\tilde{\beta}_{j,t-1}\right)-\frac{\eta^2}{V}\tilde{\beta}_{j,t}\\
&\qquad\qquad=\frac{\phi}{W}w_{(j-1)n+t+1}-\frac{1}{W}w_{(j-1)n+t}+\frac{\eta}{V}v_{(j-1)n+t}\\
&\qquad\qquad\text{for } t=n-1,\dots,1\\
&\frac{\phi}{W}\left(\tilde{\beta}_{j,1}-\phi\tilde{\beta}_{j,0}\right)-\left(A-\frac{C^2}{G^*_{(j-1)n}(D)}\right)\tilde{\beta}_{j,0}\\
&\qquad\qquad=\frac{\phi}{W}w_{(j-1)n+1}-\left(A-\frac{C^2}{G^*_{(j-1)n}(D)}\right)\tilde{\beta}_{j-1,n}
\end{aligned}
$$

and has the following distribution

$$
\boldsymbol{M}^s_j\tilde{\boldsymbol{\beta}}^s_j\sim\mathrm{N}\left(\mathbf{0},\boldsymbol{M}^s_j\right)\ \text{ or }\ \tilde{\boldsymbol{\beta}}^s_j\sim\mathrm{N}\left(\mathbf{0},\left(\boldsymbol{M}^s_j\right)^{-1}\right)
$$

$$
\boldsymbol{M}^s_j\equiv\begin{bmatrix}
A & -C & & & \\
-C & B & -C & & \\
 & & \ddots & & \\
 & & -C & B & -C\\
 & & & -C & B-\frac{C^2}{G^*_{(j-1)n}(D)}
\end{bmatrix}
$$

where the precision for $\hat{\beta}^s_{j,n}$, as shown in Lemma 3.12, is

$$\text{Var }\tilde{\beta}^s_{j,n} = \left(A - \frac{C^2}{G^*_{jn}(D)}\right)^{-1}.$$

Precision formulas for each of the partial smoothers are easy to find using the $\boldsymbol{M}^s_j$ matrices for $j = 1, \ldots, r$.

Lemma 3.12. *Given the linear Gaussian state space model (91) with $N = rn$, then precisions for the partial smoothers in $\hat{\boldsymbol{\beta}}^s_{r:1} = (\hat{\beta}^s_{j,i} : j = 1, \ldots, r;\ i = 0, \ldots, n)'$ are calculated as follows*

$$\begin{aligned}
Var\ \tilde{\beta}^s_{j,i} &= \left(G^*_{(j-1)n+i+1}(D) - \frac{C^2}{G_{n-i}}\right)^{-1},\ i = 0, \ldots, n-1 \\
Var\ \tilde{\beta}^s_{j,n} &= \left(A - \frac{C^2}{G^*_{jn}(D)}\right)^{-1}.
\end{aligned}$$

Proof: Let $\boldsymbol{X}_{n+1} \equiv (x_n, \ldots, x_0)'$. Gaussian elimination of $\boldsymbol{M}^s_j \boldsymbol{X}_{n+1} = \boldsymbol{e}_{n+1}$ shows

$$\text{Var }\tilde{\beta}^s_{j,0} = x_0 = \left(G^*_{(j-1)n+1}(D) - \frac{C^2}{G_n}\right)^{-1}$$

which proves the result for Var $\tilde{\beta}^s_{j,0}$. Gaussian elimination of $\boldsymbol{M}^s_j \boldsymbol{X}_{n+1} = \boldsymbol{e}_{n+1-i}$ where $i = 1, \ldots, n-1$ shows

$$\begin{aligned}
\text{Var }\tilde{\beta}^s_{j,i} = x_i &= \left(B - \frac{C^2}{G^*_i\left(G^*_{(j-1)n+1}(D)\right)} - \frac{C^2}{G_{n-i}}\right)^{-1} \\
&= \left(B - \frac{C^2}{G^*_{(j-1)n+i}(D)} - \frac{C^2}{G_{n-i}}\right)^{-1}
\end{aligned}$$

which is equivalent to the result for Var $\tilde{\beta}^s_{j,i}$, $i = 1, \ldots, n-1$. Gaussian elimination of $\boldsymbol{M}^s_j \boldsymbol{X}_{n+1} = \boldsymbol{e}_1$ shows

$$\begin{aligned}
\text{Var }\tilde{\beta}^s_{j,n} = x_n &= \left(A - \frac{C^2}{G^*_n\left(G^*_{(j-1)n+1}(D)\right)}\right)^{-1} \\
&= \left(A - \frac{C^2}{G^*_{jn}(D)}\right)^{-1}
\end{aligned}$$

which proves the result for Var $\tilde{\beta}^s_{j,n}$. Hence the complete result is proven. ■

Comparison of the formulas from Lemma 3.12 for the partial smoother precisions, together with the inequality $A \leq G_k$ for $k = 1, \ldots, n$, leads to the following result.

Corollary 3.8. *Given the linear Gaussian state space model (91) with $N = rn$, then for $j = 2, \ldots, r$ the precisions of $\hat{\beta}^s_{j,0}$ are better than the precisions of $\hat{\beta}^s_{j-1,n}$ where $\hat{\beta}^s_{j,0}$ and $\hat{\beta}^s_{j-1,n}$ are both partial smoother estimates of the state $\beta_{(j-1)n}$*

$$Var\ \tilde{\beta}^s_{j,0} < Var\ \tilde{\beta}^s_{j-1,n}\ .\ ■$$

Comparison of the formulas, from Lemma 3.3 for the Kalman filter precisions and from Lemma 3.12 for the partial smoother precisions, shows the following result.

Corollary 3.9. *Given the linear Gaussian state space model (91) with $N = rn$, then for $j = 1, \ldots, r$ the Kalman filter estimates $\beta_{jn|jn}$ and the partial smoother estimates $\hat{\beta}^s_{jn}$ of the states β_{jn} have the same precision*

$$Var\ \tilde{\beta}_{jn|jn} = Var\ \tilde{\beta}^s_{jn}\ .\ ■$$

The asymptotic limits on the precisions associated with the most recent partition of partial smoothers estimates $\hat{\boldsymbol{\beta}}^s_r$ are found by using the limit property of $G^*_r(D)$ from Lemma 3.6.

Proposition 3.12. *Given the linear Gaussian state space model (91) with $N = rn$, then precisions of the partial smoother estimates in the rth partition $\hat{\boldsymbol{\beta}}^s_r$ converge as $r \to \infty$*

$$\begin{aligned} Var\ \tilde{\beta}^s_{r,i} &\to \left(G^*_\infty - \frac{C^2}{G_{n-i}}\right)^{-1},\ i = 0, \ldots, n-1 \\ Var\ \tilde{\beta}^s_{r,n} &\to \left(A - \frac{C^2}{G^*_\infty}\right)^{-1}\ .\ ■ \end{aligned}$$

The next result of this section relates the precisions of the partial smoother estimates $\hat{\boldsymbol{\beta}}^s_{0:rn} \equiv \{\hat{\beta}^s_t : t = 0, \ldots, rn\}$ to the precisions of the state space smoother estimates $\boldsymbol{\beta}^k_{0:rn|rn} \equiv \{\beta_{t|rn} : t = 0, \ldots, rn\}$ and to the precisions of the other partial smoother estimates $\hat{\boldsymbol{\beta}}^m_{0:rn} \equiv \{\hat{\beta}^m_{t|rn} : t = 0, \ldots, rn\}$ where the partitions sizes associated with $\hat{\boldsymbol{\beta}}^m_{0:rn}$ are chosen such that the first element in each partition $\hat{\boldsymbol{\beta}}^m_j = (\hat{\beta}^m_{j,i} : i = n_j, \ldots, 1)'$ and $\hat{\boldsymbol{\beta}}^s_j = (\hat{\beta}^s_{j,i} : i = n, \ldots, 0)'$ are estimating the same state β_{jn} for $j = 1, \ldots, r$

$$
\begin{gathered}
n_1 = n+1,\ n_2 = \cdots = n_r = n \\
\boldsymbol{M}^m_1 \in \mathbb{R}^{n+1 \times n+1},\ \boldsymbol{M}^m_2 = \cdots = \boldsymbol{M}^m_r \in \mathbb{R}^{n \times n} \ .
\end{gathered}
$$

Theorem 3.3. *Given the linear Gaussian state space model (91) with $N = rn$, then the precisions of the state space smoothers $\beta_{t|rn} \in \boldsymbol{\beta}^k_{0:rn|rn}$ and of the partial smoothers $\hat{\beta}^m_t \in \hat{\boldsymbol{\beta}}^m_{0:rn}$ and $\hat{\beta}^s_t \in \hat{\boldsymbol{\beta}}^s_{0:rn}$ are related as follows*

$$
\begin{aligned}
&Var\ \tilde{\beta}_{0|rn} \leq Var\ \tilde{\beta}^m_{0|rn} \leq Var\ \tilde{\beta}^s_0 \\
&Var\ \tilde{\beta}_{t|rn} \leq Var\ \tilde{\beta}^s_t \leq Var\ \tilde{\beta}^m_{t|rn} \ \textit{for}\ t = 1, \ldots, rn \ .
\end{aligned}
$$

Proof: With regards to the lower bound, applying the linear Gaussian state space model (91) to the residual of the partial smoother estimate $\hat{\beta}^s_0$ of the initial state β_0 shows

$$
\tilde{\beta}^s_0 \equiv \beta_0 - \hat{\beta}^s_0 \equiv \beta_0 - b_0 \sim \mathrm{N}\left(0, W_0\right) \ .
$$

The combination of the previous display together with the precision formulas at $t = 0$ for the complete state space smoother from Lemma 3.10 and for the partial smoothers from Propositions 3.10 and 3.11 are applied to show the full result at $t = 0$

$$
\mathrm{Var}\ \tilde{\beta}_{0|rn} \leq \mathrm{Var}\ \tilde{\beta}^m_{0|rn} \leq \mathrm{Var}\ \tilde{\beta}^l_{0|0} = W_0 = \mathrm{Var}\ \tilde{\beta}^s_0 \ .
$$

Direct comparison of the precision formulas for the complete state space smoothers from Lemmas 3.3 and 3.4 for $\beta_{t|rn} = \beta_{(j-1)n+i|rn}$ and for the partial smoothers from Lemma 3.12 for $\hat{\beta}^s_t = \hat{\beta}_{(j-1)n+i} = \hat{\beta}^s_{j,i}$ where $j = 1, \ldots, r$ and $i = 1, \ldots, n$ shows

$$
\mathrm{Var}\ \tilde{\beta}_{t|rn} \leq \mathrm{Var}\ \tilde{\beta}^s_t \ \text{for}\ t = 1, \ldots, rn \ .
$$

The combination of the two previous displays proves the result for the lower bound.

With regards to the upper bound, the result has already been proven for $t = 0$. Precision formulas of the partial smoothers from Proposition 3.11 for $\hat{\boldsymbol{\beta}}^m_1$ and from the introduction to this section for $\hat{\boldsymbol{\beta}}^s_1$ are compared to show

$$
\mathrm{Var}\ \tilde{\boldsymbol{\beta}}^m_1 = (\boldsymbol{M}^m_1)^{-1} = \boldsymbol{M}^{-1}_n = (\boldsymbol{M}^s_1)^{-1} = \mathrm{Var}\ \tilde{\boldsymbol{\beta}}^s_1
$$

such that

$$
\begin{aligned}
\text{Var } \tilde{\beta}^m_{1,i} &= \text{Var } \tilde{\beta}^s_{1,i} \text{ for } i = 0, \dots, n \\
\text{Var } \tilde{\beta}^m_{t|rn} &= \text{Var } \tilde{\beta}^s_t \text{ for } t = 1, \dots, n \ .
\end{aligned} \tag{124}
$$

The system of equations for $\tilde{\boldsymbol{\beta}}^s_j$ for $j = 2, \dots, r$ is rewritten to show

$$
\begin{bmatrix} A & -C & & \\ -C & B & -C & \\ & & \ddots & \\ & & -C & B \end{bmatrix} \begin{pmatrix} \tilde{\beta}^s_{j,n} \\ \tilde{\beta}^s_{j,n-1} \\ \vdots \\ \tilde{\beta}^s_{j,1} \end{pmatrix}
$$
$$
= \begin{pmatrix} \frac{1}{W} w_{jn} - \frac{\eta}{V} v_{jn} \\ -\frac{\phi}{W} w_{jn} + \frac{1}{W} w_{(j-1)n+n-1} - \frac{\eta}{V} v_{(j-1)n+n-1} \\ \vdots \\ -\frac{\phi}{W} w_{(j-1)n+2} + \frac{1}{W} w_{(j-1)n+1} - \frac{\eta}{V} v_{(j-1)n+1} + C\tilde{\beta}^s_{j,0} \end{pmatrix}
$$

which leads directly to the precision for the partial smoothers $\hat{\boldsymbol{\beta}}^s_{j,n:1} = (\hat{\beta}_{j,i} : i = n, \dots, 1)$

$$
\begin{aligned}
\text{Var}\left(\boldsymbol{M}^m_j \tilde{\boldsymbol{\beta}}^s_{j,n:1}\right) &= \boldsymbol{M}^m_j + C^2 \text{Var}\left(\tilde{\beta}^s_{j,0}\right) \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 1 \end{bmatrix} \\
\text{Var}\left(\tilde{\boldsymbol{\beta}}^s_{j,n:1}\right) &= \left(\boldsymbol{M}^m_j\right)^{-1} + C^2 \text{Var}\left(\tilde{\beta}^s_{j,0}\right) \boldsymbol{Z}_n \boldsymbol{Z}'_n \\
\left(\boldsymbol{M}^m_j\right)^{-1} &= [z_{i_1,i_2} : i_1, i_2 = 1, \dots, n] \\
\boldsymbol{Z}_n &= (z_{i,n} : i = 1, \dots, n)' \ .
\end{aligned}
$$

Proposition 3.11 also provides the precision for the partial smoothers $\hat{\boldsymbol{\beta}}^m_j$, $j = 2, \dots, r$

$$
\text{Var}\left(\tilde{\boldsymbol{\beta}}^m_j\right) = \left(\boldsymbol{M}^m_j\right)^{-1} + C^2 \text{Var}\left(\tilde{\beta}^m_{j-1,n}\right) \boldsymbol{Z}_n \boldsymbol{Z}'_n \ .
$$

For $j = 2$, applying (124) and Corollary 3.8 to the previous two displays shows

$$
\begin{aligned}
\text{Var } \tilde{\beta}^s_{2,i} &\le \text{Var } \tilde{\beta}^m_{2,i} \text{ for } i = 1, \dots, n \\
\text{Var } \tilde{\beta}^s_t &\le \text{Var } \tilde{\beta}^m_{t|rn} \text{ for } t = n+1, \dots, 2n \ .
\end{aligned}
$$

Induction for $j = 3, \ldots, r$ is used to complete the proof of the result for the upper bound. Hence the result is proven. ■

With the partition size set to $n = 1$, Lemma 3.12 shows that the precisions for the partial smoother estimates $\hat{\boldsymbol{\beta}}^{s}_{0:N} = \{\hat{\beta}^s_t : t = 0, \ldots, N\}$ are the same as the precisions for the Kalman filter estimates $\boldsymbol{\beta}^{t|t}_{0:N} = \{\beta_{t|t} : t = 0, \ldots, N\}$. The next result in this section shows when $n = 1$ that in fact the partial smoother estimates $\hat{\boldsymbol{\beta}}^{s}_{0:N}$ are equivalent to the Kalman filter estimates $\boldsymbol{\beta}^{t|t}_{0:N}$ and also shows that the initial partial smoother estimates $\hat{\boldsymbol{\beta}}^{s+1}_{0:N} = \{\hat{\beta}^s_{t,0} : t = 1, \ldots, N\}$ are equivalent to the one step state space smoother estimates $\boldsymbol{\beta}^{t-1|t}_{0:N} = \{\beta_{t-1|t} : t = 1, \ldots, N\}$.

Theorem 3.4. *Given the linear Gaussian state space model (91) with $N = r$ and given $n = 1$, then the partial smoother estimates $\hat{\boldsymbol{\beta}}^{s}_{0:N}$ and the Kalman filter estimates $\boldsymbol{\beta}^{t|t}_{0:N}$ are equivalent*

$$\hat{\beta}^s_t = \boldsymbol{\beta}_{t|t} \textit{ for } t = 0, \ldots, N$$

and the initial partial smoother estimates $\hat{\boldsymbol{\beta}}^{s+1}_{0:N}$ and the one step state space smoother estimates $\boldsymbol{\beta}^{t-1|t}_{0:N}$ are equivalent

$$\hat{\beta}^s_{t,0} = \boldsymbol{\beta}_{t-1|t} \textit{ for } t = 1, \ldots, N\ .$$

Proof: Proposition 3.3 proved the following results

$$\begin{aligned}
\beta_{0|0} &\equiv b_0 \\
\beta_{1|1} &= \left(A - \frac{C^2}{G^*_1(D)}\right)^{-1} \left(\frac{\eta}{V} Y_1 + \frac{C}{G^*_1(D)} \frac{b_0}{W_0}\right) \\
\beta_{2|2} &= \left(A - \frac{C^2}{G^*_2(D)}\right)^{-1} \left(\frac{\eta}{V} Y_2 + \frac{C}{G^*_2(D)} \left(\frac{\eta}{V} Y_1 + \frac{C}{G^*_1(D)} \frac{b_0}{W_0}\right)\right) \\
\beta_{t|t} &= \left(A - \frac{C^2}{G^*_t(D)}\right)^{-1} \left(\frac{\eta}{V} Y_t + \frac{C}{G^*_t(D)} \left(A - \frac{C^2}{G^*_{t-1}(D)}\right) \beta_{t-1|t-1}\right) \\
&\quad \text{for } t = 2, \ldots, N\ .
\end{aligned}$$

The system of state estimating equations for the partial smoother estimates with $n = 1$ at time indices $0, 1, 2, t$ are

$$\begin{aligned}
\hat{\beta}^s_0 &\equiv b_0 \\
\begin{bmatrix} A & -C \\ -C & G^*_1(D) \end{bmatrix} \begin{pmatrix} \hat{\beta}^s_{1,1} \\ \hat{\beta}^s_{1,0} \end{pmatrix} &= \begin{pmatrix} \frac{\eta}{V} Y_1 \\ \frac{b_0}{W_0} \end{pmatrix} \\
\begin{bmatrix} A & -C \\ -C & G^*_2(D) \end{bmatrix} \begin{pmatrix} \hat{\beta}^s_{2,1} \\ \hat{\beta}^s_{2,0} \end{pmatrix} &= \begin{pmatrix} \frac{\eta}{V} Y_2 \\ \left(A - \frac{C^2}{G^*_1(D)} \right) \beta^s_{1,1} \end{pmatrix} \\
\begin{bmatrix} A & -C \\ -C & G^*_t(D) \end{bmatrix} \begin{pmatrix} \hat{\beta}^s_{t,1} \\ \hat{\beta}^s_{t,0} \end{pmatrix} &= \begin{pmatrix} \frac{\eta}{V} Y_t \\ \left(A - \frac{C^2}{G^*_{t-1}(D)} \right) \beta^s_{t-1,1} \end{pmatrix} \ .
\end{aligned}$$

Gaussian elimination of each system of state estimating equations in the previous display to remove the upper diagonal in each (2×2) matrix in order to solve for $\hat{\beta}^s_{j,1}$ for $j = 1, 2, t$ shows that the partial smoother estimates with $n = 1$ are equivalent to the Kalman filter estimates

$$\beta_{0|0} = \hat{\beta}^s_0 \text{ and } \beta_{j|j} = \hat{\beta}^s_{j,1} = \hat{\beta}^s_j \text{ for } j = 1, 2, t \ .$$

Hence the first result for the Kalman filter estimates is proven by induction.

The system of equations that the state space smoother estimates satisfy, $\boldsymbol{M}_N \boldsymbol{\beta}^k_{N:0} = \boldsymbol{Y}^*_{N:0}$ with $N = t$, shows that the Kalman filter $\beta_{t|t}$ and the one step smoother $\beta_{t-1|t}$ are related as follows

$$A\beta_{t|t} - C\beta_{t-1|t} = \frac{\eta}{V} Y_t \text{ for } t = 1, \dots, N \ .$$

The corresponding system of equations for the tth partition of partial smoother estimates $\hat{\boldsymbol{\beta}}^s_t \equiv \{\hat{\beta}^s_{t,1}, \hat{\beta}^s_{t,0}\}$ shows the following relationship

$$A\hat{\beta}^s_{t,1} - C\hat{\beta}^s_{t,0} = \frac{\eta}{V} Y_t \text{ for } t = 1, \dots, N \ .$$

The first result of this lemma and the two previous displays prove the second result

$$\beta_{t-1|t} = \hat{\beta}^s_{t,0} \text{ for } t = 1, \dots, N \ . \blacksquare$$

With the partition size set to $n \geq 1$, the final result in this section generalizes the result from the previous lemma to show how the partial smoother estimates and the state space smoother estimates of each state are related.

Theorem 3.5. *Given the linear Gaussian state space model (91) with $N = rn$, then the partial smoother partitions $\hat{\boldsymbol{\beta}}_j^s$ and the state space smoother estimates $\boldsymbol{\beta}^k_{(j-1)n:jn|jn}$ are equivalent for $j = 1, \ldots, r$*

$$\hat{\beta}^s_{j,i} = \beta_{(j-1)n+i|nj} \textit{ for } j = 1, \ldots, r;\ i = 0, \ldots, n\ .$$

Proof: With respect to the first partition, the partial smoother $\hat{\boldsymbol{\beta}}_1^s$ and the corresponding state space smoother $\beta^k_{n:0|n}$ both satisfy the same system of equations

$$\boldsymbol{M}_1^s \hat{\boldsymbol{\beta}}_1^s = \boldsymbol{Y}^*_{n:0},\ \boldsymbol{M}_n \boldsymbol{\beta}^k_{n:0|n} = \boldsymbol{Y}^*_{n:0},\ \boldsymbol{M}_1^s = \boldsymbol{M}_n\ .$$

Hence the result is proven for the first partition since $\boldsymbol{M}_n$ is invertible and the two solutions are equivalent

$$\hat{\boldsymbol{\beta}}_1^s = \boldsymbol{\beta}^k_{n:0|n}\ .$$

Using Gaussian elimination to solve for $\hat{\beta}^s_{1,n} = \beta_{n|n}$ by eliminating the upper diagonal in $\boldsymbol{M}_n$ shows that the solution is

$$\begin{aligned} \beta_{n|n} = {} & \left(A - \frac{C^2}{G_n^*(D)}\right)^{-1} \left(\frac{\eta}{V} Y_n + \frac{C}{G_n^*(D)} \left(\frac{\eta}{V} Y_{n-1} + \ldots \right.\right. \\ & \left.\left. + \frac{C}{G_2^*(D)} \left(\frac{\eta}{V} Y_1 + \frac{C}{G_1^*(D)} \frac{b_0}{W_0}\right) \ldots \right)\right)\ . \end{aligned}$$

With respect to the second partition, the partial smoother $\hat{\boldsymbol{\beta}}_2^s$ satisfies the following system of equations

$$\begin{bmatrix} A & -C & & & \\ -C & B & -C & & \\ & & \ddots & & \\ & & -C & B & -C \\ & & & -C & G^*_{n+1}(D) \end{bmatrix} \begin{pmatrix} \hat{\beta}^s_{2,n} \\ \hat{\beta}^s_{2,n-1} \\ \vdots \\ \hat{\beta}^s_{2,1} \\ \hat{\beta}^s_{2,0} \end{pmatrix} = \begin{pmatrix} \frac{\eta}{V} Y_{2n} \\ \frac{\eta}{V} Y_{2n-1} \\ \vdots \\ \frac{\eta}{V} Y_{n+1} \\ \left(A - \frac{C^2}{G_n^*(D)}\right) \hat{\beta}^s_{1,n} \end{pmatrix}\ .$$

Gaussian elimination of the system of equations in the previous display to solve for $\hat{\beta}^s_{2,n}$ by eliminating the upper diagonal in the square matrix shows that the solution is

$$
\begin{aligned}
\hat{\beta}^s_{2,n} &= \left(A - \frac{C^2}{G^*_{2n}(D)}\right)^{-1} \left(\frac{\eta}{V} Y_{2n} + \frac{C}{G^*_{2n}(D)} \left(\frac{\eta}{V} Y_{2n-1} + \ldots \right.\right. \\
&\left.\left. + \frac{C}{G^*_{n+2}(D)} \left(\frac{\eta}{V} Y_{n+1} + \frac{C}{G^*_{n+1}(D)} \left(A - \frac{C^2}{G^*_n(D)}\right) \hat{\beta}^s_{1,n}\right) \ldots \right)\right) \\
&= \left(A - \frac{C^2}{G^*_{2n}(D)}\right)^{-1} \left(\frac{\eta}{V} Y_{2n} + \frac{C}{G^*_{2n}(D)} \left(\frac{\eta}{V} Y_{2n-1} + \ldots \right.\right. \\
&\left.\left. + \frac{C}{G^*_2(D)} \left(\frac{\eta}{V} Y_1 + \frac{C}{G^*_1(D)} \frac{b_0}{W_0}\right) \ldots \right)\right) .
\end{aligned}
$$

Hence $\hat{\beta}^s_{2,n} = \beta_{2n|2n}$ since $\hat{\beta}^s_{2,n}$ has the same solution as $\beta_{2n|2n}$ where $\beta_{2n|2n}$ is found by Gaussian elimination of $\boldsymbol{M}_{2n}\boldsymbol{\beta}^k_{2n:0|2n} = \boldsymbol{Y}^*_{2n:0}$. The system of equations associated with the partial smoothers $\{\hat{\beta}^s_{2,1}, \ldots, \hat{\beta}^s_{2,n}\}$ and the state space smoothers $\boldsymbol{\beta}^k_{n+1:2n|2n}$ shows that both sets of smoothers satisfy the same system of equations

$$
\begin{aligned}
A\hat{\beta}^s_{2,n} - C\hat{\beta}^s_{2,n-1} &= \frac{\eta}{V} Y_{2n} \\
A\beta_{2n|2n} - C\beta_{2n-1|2n} &= \frac{\eta}{V} Y_{2n}
\end{aligned}
$$

and for $i = n-1, \ldots, 1$

$$
\begin{aligned}
-C\hat{\beta}^s_{2,i+1} + B\hat{\beta}^s_{2,i} - C\hat{\beta}^s_{2,i-1} &= \frac{\eta}{V} Y_{n+i} \\
-C\beta_{n+i+1|2n} + B\beta_{n+i|2n} - C\beta_{n+i-1|2n} &= \frac{\eta}{V} Y_{n+i} .
\end{aligned}
$$

Hence the result is proven for the second partition

$$
\hat{\beta}^s_{2,i} = \beta_{n+i|2n} \text{ for } i = n, \ldots, 0 .
$$

Induction is used to prove the result for the remaining partitions

$$
\hat{\beta}^s_{j,i} = \beta^k_{(j-1)n+i|jn} \text{ for } j = 3, \ldots, r;\; i = n, \ldots, 0 .
$$

Hence the complete result is proven. ■

References

[1] Amemiya, T. (1985). *Advanced Economics.* Harvard University Press, Cambridge.

[2] Bickel, P. J., and K. A. Doksum (1977). *Mathematical Statistics.* Prentice Hall, New Jersey.

[3] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994). *Time Series Analysis Forecasting and Control, Third Edition.* Prentice Hall, New Jersey.

[4] Brockwell, P. J., and R. A. Davis (1991). *Time Series: Theory and Methods, Second Edition.* Springer, New York.

[5] Bryson A. E., and M. Frazier (1963). Smoothing for linear and nonlinear dynamic systems. *TDR 63-119, Tech. Rep.*, 353-364. Aero. Sys. Div. Wright-Patterson Air Force Base, Ohio.

[6] Casella, G., and R. Berger (1990). *Statistical Inference.* Duxbury Press, Belmont.

[7] Durbin, J., and S. J. Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford.

[8] Fokianos, K., B. Kedem, J. Qin, J. Haferman, and D. A. Short (1998). On Combining Instruments. *Journal of Applied Meteorology*, **37**, 220-226.

[9] Fokianos, K., B. Kedem, J. Qin, and D. A. Short (2001). A semiparametric approach to the one-way layout. *Technometrics*, **43**, 56-65.

[10] Gelb, A. J. F., Kasper, R. A. Nash, C. F. Price, and A. A. Sutherland (1974). *Applied Optimal Estimation.* M.I.T. Press, Cambridge.

[11] Gilbert, P. B., S. R. Lele, and Y. Vardi (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, **86**, 27-43.

[12] Gilber, P. B. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling models. *Annals of Statistics*, **28**, 151-194.

[13] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASME. Series D*, **82**, 35-45.

[14] Kedem, B., and K. Fokianos (2002). *Regression Models for Time Series Analysis.* Wiley, New York.

[15] Kedem, B., D. B. Wolff, and K. Fokianos (2004). Statistical Comparison of Algorithms. *IEEE Trans. Instrum. Meas.*, **53**, 770-776.

[16] Kress, R. (1998). *Numerical Analysis*. Springer-Verlag, New York.

[17] Künsch, H. R. (2001). State space and Hidden Markov models. In E. Barndorff-Nielsen et al., editors, *Complex Stochastic Systems,* 109-173, Chapman & Hall/CRC, Boca Raton, Florida.

[18] Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, New York.

[19] Patil, G. P., and C. R. Rao (1977), The weighted distributions: A survey of their applications. *Applications of Statistics*, Ed. P.R. Krishnaiah, 383-405. Amsterdam: North-Holland.

[20] Qin, J. (1998), Inferences for case-control and semiparametric two-sample density ratio models. *Biometrika*, **85**, 619-630.

[21] Qin, J., and B. Zhang (1997), A Goodness of fit test for logistic regression models based on case-control data. *Biometrika*, **84**, 609-618.

[22] Petrov, V. V. (1995). *Limit Theorems of Probability Theory.* Oxford University Press, New York.

[23] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[24] Rauch, H. E. (1962). Linear estimation of sampled stochastic processes with random parameters. *Tech. Rep. 2108-1.* Stanford Electronics Laboratory, Stanford University, Stanford, CA.

[25] Rauch H. E., F. Tung, and C. T. Striebel (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal*, **3**, 1445-1450.

[26] Rhodes, I. B. (1971). A tutorial introduction to estimation and filtering. *IEEE Transactions on Automatic Control*, **16**, 688-706.

[27] Scheffe, H. (1959). *The Analysis of Variance.* Wiley, New York.

[28] Seng, P. K., and J. M. Singer (1993), *Large Sample Methods in Statistics, An Introduction With Applications*, Chapman and Hall, London.

[29] Shumway, R. H., and D. S. Stoffer (2000). *Time Series Analysis and Its Applications.* Springer, New York.

[30] van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.